

5-13-2011

The Phylogeography of *Marstonia Lustrica*: Understanding the Relationship Between Glaciation and the Evolution and Distribution of a Rare Snail

Thomas W. Coote

University of Massachusetts Amherst, tcoote@eco.umass.edu

Follow this and additional works at: https://scholarworks.umass.edu/open_access_dissertations



Part of the [Agricultural and Resource Economics Commons](#)

Recommended Citation

Coote, Thomas W., "The Phylogeography of *Marstonia Lustrica*: Understanding the Relationship Between Glaciation and the Evolution and Distribution of a Rare Snail" (2011). *Open Access Dissertations*. 399.

https://scholarworks.umass.edu/open_access_dissertations/399

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

THE PHYLOGEOGRAPHY OF *MARSTONIA LUSTRICA*:
UNDERSTANDING THE RELATIONSHIP BETWEEN GLACIATION
AND THE EVOLUTION AND DISTRIBUTION OF A RARE SNAIL.

A Dissertation Presented

By

THOMAS W. COOTE

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2011

Department of Environmental Conservation

THE PHYLOGEOGRAPHY OF *MARSTONIA LUSTRICA*:
UNDERSTANDING THE RELATIONSHIP BETWEEN GLACIATION
AND THE EVOLUTION AND DISTRIBUTION OF A RARE SNAIL.

A Dissertation Presented

By

THOMAS W. COOTE

Approved as to style and content by:

Kevin McGarigal, Chair

Andrew Whiteley, Member

Douglas G. Smith, Member

Stephen DeStephano, Member

Donald Roeder, Member

Robert Schmidt, Member

Paul R. Fisette, Department Head
Environmental Conservation

ABSTRACT

THE PHYLOGEOGRAPHY OF *MARSTONIA LUSTRICA*: UNDERSTANDING THE RELATIONSHIP BETWEEN GLACIATION AND THE EVOLUTION AND DISTRIBUTION OF A RARE SNAIL.

MAY 2011

THOMAS W. COOTE, PH.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Kevin McGarigal

Marstonia lustrica is a poorly understood aquatic snail, relatively rare throughout its range and listed in the State of Massachusetts as Endangered (MNHESP 2010, Hershler et. al 1987). It is the northern-most cold temperate species of its genus, with other members of the genus occurring along the southern edge of its range and in the southeastern United States (Thompson 1977). The current range of *M. lustrica* appears to follow the maximum extent of the Laurentide Glacier (20–25 kya), extending from Minnesota to western Massachusetts. Research regarding the distribution, ecology, and phylogeny of *M. lustrica* in the State of Massachusetts and eastern New York raised the possible role of glaciers and pro-glacial lakes in the establishment and distribution of the snail, leading to the hypothesis that its distribution and evolution may be dependent upon glacial processes. A full range survey was completed in 2007 and 2008, with populations identified in 20 water bodies from Minnesota to Massachusetts, and Ohio to Ontario, Canada. Fifty-seven specimens from the 20 populations were sequenced for two mtDNA markers (COI and NDI), developing both phylogenetic trees and haplotype networks. Here I present those trees and networks, and correlate the distribution of these populations and their representative haplotypes with both glacial events and contemporary watersheds, using AMOVAs and Mantel tests to examine several

phylogeographic models. In addition to the results for *M. lustrica*, the unexpected occurrence of several other species of *Marstonia* spp. found across the range of *M. lustrica* are presented, including *M. pachyta*, *M. comalensis*, and *M. hershleri*.

TABLE OF CONTENTS

ABSTRACT.....	iv
LIST OF TABLES	viii
LIST OF FIGURES.....	ix
CHAPTER	
1. INTRODUCTION.....	1
Background.....	4
Objectives.....	7
2. LITERATURE REVIEW.....	11
Ecology.....	11
Taxonomy and Morphology.....	15
Biogeography across the Range.....	17
Biogeography of the Eastern Populations.....	20
Glaciation.....	25
Taxonomy and Genetics.....	34
3. METHODOLOGY	39
Fieldwork Protocol.....	40
Taxonomy and Morphology.....	43
Phylogenetics.....	45
Extraction, PCR, and Sequencing	45
Quantitative Analysis.....	48
4. RESULTS.....	51
Morphology.....	51
Phylogenetics.....	60
Sequence Analysis.....	62
Quantitative Analysis.....	74
5. DISCUSSION.....	79

6. CONCLUSION.....	87
Conservation.....	93
APPENDICES	
A. LIST OF LAKES, RIVERS AND PONDS SAMPLED IN 2000-05.....	96
B. LIST OF LAKES, RIVERS AND PONDS SAMPLED IN 2007-08.....	97
C. COI SEQUENCE TEXT FILES.....	98
D. NDI SEQUENCE TEXT FILES.....	109
E. LIST OF SEQUENCED POPULATIONS	115
F. EXTRACTION PROTOCOL	116
G. GLACIAL MAPS.....	117
H. NUCLEOTIDE SUBSTITUTION PATTERNS.....	125
I. HAPLOTYPE TABLE FOR COI.....	126
J. HAPLOTYPE TABLE FOR NDI.....	127
LITERATURE CITED.....	128

LIST OF TABLES

Table	Page
1. List of other <i>Marstonia</i> species.....	21
2. Glacial events and time frames.....	28
3. Major glacial lakes, time frames, and flow regimes.....	31
4. List of morphological character states.....	44
5. Regional cluster assignments for AMOVAs.....	50
6. List of lakes and water quality results.....	52
7. Haplotype list from pilot study.....	61
8. AMOVA results for COI sequences.....	75
9. Mantel results for COI and Euclidian Distance.....	76
10. Mantel results for NDI and Euclidian Distance.....	76
11. Mantel results for Watershed and Cell.....	77
12. Partial Mantel results for Watershed and Cell.....	77
13. AMOVAs for COI and NDI.....	78

LIST OF FIGURES

Figure	Page
1. Historical populations of <i>Marstonia lustrica</i>	2
2. Map of other <i>Marstonia</i> species.....	19
3. Map of Glacial Lake Bascom.....	26
4. Map of major glaciations.....	29
5. Map of cells.....	41
6. Nomenclature for verge illustrations.....	54
7. Verges from select populations.....	55
8. Verges from select <i>Marstonia</i> sp.....	56
9. Verges from select <i>Marstonia</i> sp.....	57
10. Shells from select <i>Marstonia</i> sp.....	58
11. Shells from select populations.....	59
12. Neighbor-joining tree for COI.....	63
13. Neighbor-joining tree for NDI.....	64
14. Neighbor-joining tree for concatenated data.....	65
15. Collapsed tree for COI with outspecies.....	66
16. Haplotype network for COI.....	69
17. Haplotype network for NDI.....	70
18. Haplotype network for concatenated data.....	71
19. Phylogenetic tree of concatenated data with hypothesized species associations..	72
20. Map of sequenced populations and common haplotypes.....	80
21. Map of hypothesized species relationships and divergent events.....	85
22. Phylogenetic tree with associated glacial time periods.....	86

CHAPTER 1

INTRODUCTION

It is well known that global climate cycles and glacier advance and retreat have played a significant role in the distribution and evolution of a number of plants and animals, but the exact process is not well understood (Schmidt 1986; Hewitt 1995, 2000 and 2004; Yang et al 2001; Rowe et al 2004; Emerson and Hewitt 2005; Curry 2006). Here I attempt to integrate the recent glacial history of North America with the distribution and population structure of a rare snail genus *Marstonia* generally, and for *M. lustrica* in particular.

Marstonia lustrica is a poorly understood aquatic snail, relatively rare throughout its range and listed in the State of Massachusetts as Endangered where it is known from only two lakes (MNHESP 2010; Ludlam et al 1973). Among the largest group of aquatic mollusks, *M. lustrica* is an operculate, prosobranch snail, part of the freshwater subfamily Nymphophilinae (Hydrobidae) containing 159 species in ten genera (Hershler et al 2003). It is the northern-most cold-temperate species of its genus, with other members of the genus occurring along the southern edge of its range (Thompson 1977), which appears to follow the maximum extent of the Laurentide Glacier (20-25 kya). It is concentrated around the Great Lakes in Michigan, with populations decreasing in the east, and occurs at relatively low frequencies at its eastern extent in New York and Massachusetts (Fig. 1) (Berry 1943; Burch 1980; Harman and Berg 1971; Thompson 1977; Jokinen 1992).

Research regarding the distribution, ecology, and phylogeny of *M. lustrica* in the State of Massachusetts and eastern New York (Wagner, pers. com.; Coote and Roeder 1999; Roeder and Coote 2000; Coote unpub. 2005-2006), raised the possible role of glaciers and pro-glacial lakes in the establishment and distribution of the snail. That research suggests that its habitat relationships are less confined than previously thought, that *M. lustrica* exhibits atypically low vagility and fecundity, and that traditional

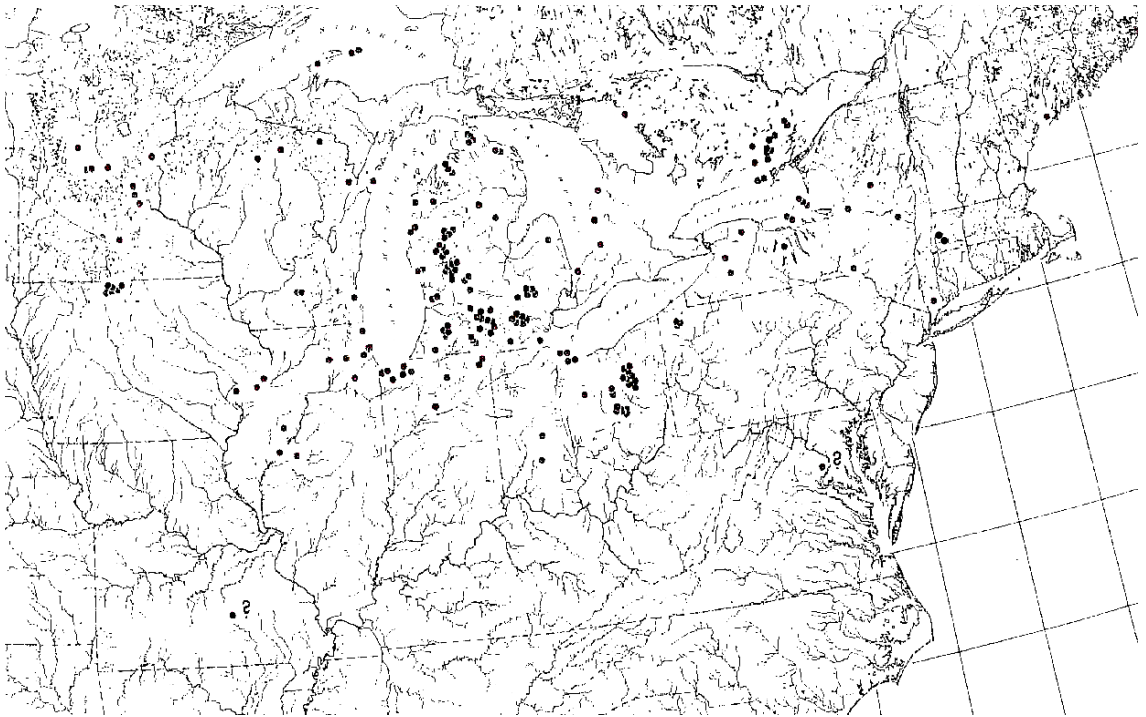


Figure 1 – Historical populations of *Marstonia lustrica*. Question marks (original) at the southern edge indicate questionable identifications (original). The two populations on the border of western Massachusetts were added here (Thompson 1977).

explanations for distribution and establishment of gastropods were unlikely to apply.

Similar to some other widespread but isolated hydrobiid species, random transport by

agents such as fish, birds, or humans, and successful establishment of populations, appears to be unlikely for this species (Hershler et al 2003 & 2008). The presence of the snail in glacial potholes in MA pointed to the possible role of glaciers in its distribution, leading to the hypothesis that it may be dependent upon forces and time scales beyond contemporary forces.

Scale in space and time is critical to understanding the structure and function of organisms across landscapes. While significant research has been dedicated towards understanding vertebrates in the context of the landscape, significantly less research has focused on invertebrates (Lydeard et al 2004). The application of genetics and landscape ecology, in the context of glaciation, may be critical to understanding the establishment and dispersal of northern invertebrates and vertebrates. Understanding the extent to which the process of glaciation and the formation of watersheds intersect and act as forces in the promotion and resistance to gene flow is essential for the conservation of multiple species including gastropods. Here I examine a species in likely decline, address the question of glacial change as a force in its evolution at multiple scales in space and time, and the implications of global warming for a species dependent on the slow pace of glaciation.

There are patterns in the distribution of *M. lustrica* that reflect the patterns of the landscape across its range. Studying these relationships through genetics, which provides fine scale data reflecting the impact of the structure and function of the landscape on *M. lustrica* over time, increases our understanding of how to conserve this species. The long-term function of glaciation may be a driving force for this species and others like it. Genetic analysis is an efficient way to address these questions, and allows us to examine *M. lustrica* across its full geographical range, without determining *a priori* distinct

populations. These relationships are critical for understanding the presence of the snail in Massachusetts, as well as the risk of extinction across its range, and for designing management strategies that take into consideration the mechanisms that drive its distribution and rarity.

Specifically, this dissertation examines the phylogeography of *M. lustrica* across the species range from Minnesota to Massachusetts and Ohio to Canada, utilizing two mitochondrial DNA markers. Working from a foundation of landscape ecology and biogeography, I combine these disciplines with the field of landscape genetics to improve our understanding of its evolutionary history across its range, its relationship to the other species in its genus, and current status for the purpose of conservation.

Background

In 1999 and 2000 I was working for the Berkshire Environmental Research Center (BERC) on a project consulting for the Stockbridge Bowl Association (SBA) and the Massachusetts Natural Heritage and Endangered Species Program (MNHESP), assessing the status of *M. lustrica* within Stockbridge Bowl and western Massachusetts in general. At the time, Stockbridge Bowl contained the only known population of *M. lustrica* in Massachusetts, and due to its status as endangered, was a significant factor in the SBA's management regime. Significant malacological work had been conducted on the lakes and streams of western Massachusetts and across the state over the years, but there was no evidence that *M. lustrica* existed in any other water body in the state. The nearest neighboring live population is located on the shores of the Hudson River, over 20

miles to the west, 280 meters lower in elevation, and in an a separate major watershed¹. The snail's presence in Massachusetts raised several intriguing questions: how did it get there, was it native or introduced, and how was it related to the other populations? At the time, it was believed that it was probably glacial in origin, or that it had been introduced, but the mechanisms could only be guessed at.

BERC identified five ecologically similar glacial lakes in western Massachusetts in an attempt to find additional populations. Targeting what was then considered to be ideal *M. lustrica* habitat, we identified natural lakes with glacial histories, sampling shore zone areas that contained *Chara* sp. beds, other submerged aquatic vegetation, and gravelly substrates. We did not find any live *M. lustrica*, but did find a few empty shells in Laurel Lake, a few miles east of Stockbridge Bowl. Because shell morphology is an imprecise method of differentiating *Marstonia* species from other gastropods, we could not be certain of this identification. However, given the lack of morphologically competing species in New England, we were confident it was *M. lustrica*. During this time I was in communication with Dr. Robert Hershler of the Smithsonian Institute and sending him the occasional specimen for anatomical analysis. As part of his work on the phylogeny of Hydrobiidae, Dr. Hershler ran genetic analysis on some of the specimens I sent him from Stockbridge Bowl (Hershler et al 2003). This analysis established the phylogenetic relationship of *M. lustrica* among the family of hydrobiids and highlighted the possibility of using genetic analysis to investigate the distribution of the species across its range, thus making it plausible to infer its phylogeography.

¹ There is an old record for *M. decepta* (*M. lustrica*) from the lower Housatonic in Connecticut, but Jokinen and Ponder (1981) concluded they are now extinct there.

Having put the project down for several years, I returned to it in 2005 and completed additional sampling in Massachusetts, New York, and Connecticut. I also re-sampled Laurel Lake and discovered several live specimens. The lack of success in finding *M. lustrica* in other lakes in the region made it clear that in order to come to some understanding of how this species moved across the landscape, what its ecology was, and what its prospects were for conservation, a much larger scale project was required. What emerged was this dissertation, examining the species across its entire range, attempting to confirm historical records, inspect museum lots, investigate habitat associations, establish a contemporary range, and run genetic analysis on each population to infer its historical distribution patterns and phylogeny, and ultimately relate these factors to the landscape and glaciation.

While traditional ecological methods were seriously limited in their ability to answer these questions, the recent advent of genetic analysis was a tool that could provide some insight. I believed that genetics made it possible to test a number of life history scenarios, and if not quantitatively determine *the* history, at least point us in the right direction for the purpose of developing a rational conservation framework.

This dissertation consists of several parts: 1) a review of *M. lustrica*'s ecology, taxonomy and morphology, and biogeography, 2) the development and analysis of its phylogeny across its range within the genus of *Marstonia*, and 3) proposed models for its relationship to the landscape, with a particular focus on glaciation. These three areas collectively inform our understanding of how *M. lustrica* evolved in relation to the landscape, which in turn informs our understanding of its relationship to other members of its genus, all of which provides guidance for conservation.

Objectives

Using a landscape ecology framework I employed two spatial scales (hereafter referred to as macro and meso) to understand *M. lustrica*'s distribution and relationship to the landscape². Several models of population genetic structure were examined for a best fit. The null model for each is "isolation by distance"; i.e., that Euclidian geographic distance best explains the observed population genetic structure. In other words, the extent to which each population is genetically different reflects its isolation by distance from the source and neighboring populations.

At the macro scale, the overarching question is, what is the relationship between the distribution of the snail across its range, the regional watersheds, glacial history, and its phylogeny? Of particular interest are the role of glacial expansion and retreat, and the role of watersheds as forces of resistance and connection in the movement of the snail then and now as inferred through genetic analysis. The null model states that population genetic structure is due to simple Euclidian geographic distance, reflecting an interpretation of glacial history whereby all current populations of *M. lustrica* represent radiation events from a geographically isolated source population in the Ohio Valley following the retreat of the Laurentide glacier. The alternative model is that a greater degree of population genetic structure is due to isolation by watersheds which evolved from pro-glacial lakes, and suggests that the current distribution of *M. lustrica* is the

² While part of the original proposal, a third Micro level analysis is not considered here due to failed attempts to obtain sequences for the ITS1 (rRNA) marker necessary for fine-scale analysis. The original objective was to examine the genetic relationships between relatively adjacent bodies of water in the Hudson (New York) and Housatonic (Massachusetts) drainages.

result of multiple source populations, reflecting multiple glacial events and multiple population advances and retreats.

At the meso scale, the focus is on the distribution pattern of *M. lustrica* among regional watersheds, specifically from southern Ontario and western New York to the Hudson River in eastern New York, into western Massachusetts. The question here is, does the genetic structure of these populations reflect human activities (such as the construction of the Erie Canal) or passive transport via other organisms such as birds, and how can the presence of the two populations in Massachusetts best be explained? The null model at the meso scale is isolation by distance and states that the genetic structure of these populations should reflect the same isolation-by-distance pattern of the macro scale.

At the meso scale an alternative model is that the eastern populations reflect random (human or avian transport) and non-random (Erie Canal) distribution events. Accordingly, these populations should contain commonly derived haplotypes and share those haplotypes with populations outside of their respective watersheds. Specifically, *M. lustrica* populations in the east should contain midwestern haplotypes derived from the radiation event following the most recent deglaciation and should not exhibit unique haplotypes.

A second alternative model at the meso scale is that the eastern populations are more closely related to each other than to a given population in the midwest, implying a secondary source population. In other words, the eastern populations should share haplotypes restricted to their respective watersheds while containing fewer haplotypes from the midwest or western populations, reflecting the effect of watershed isolation over

long time frames. In particular, the Massachusetts populations should share more haplotypes with their closest neighboring populations in eastern New York and Canada, while simultaneously exhibiting unique haplotypes, precluding recent gene exchange as the result of “bucket events”. Such a pattern could imply a refugium other than the Ohio Valley, possibly along the eastern seaboard.

Below, the models are further developed as three overarching frameworks for understanding the distribution patterns of *M. lustrica*, which are not mutually exclusive.

1. Watershed:

This model posits that population genetic structure reflects contemporary watersheds, which in turn reflect glacial history. The various haplotypes exhibited in a given watershed have derived solely by isolation from other populations along watershed boundaries. Phylogenetic analysis should show that haplotypes are structured by watershed, and analysis of molecular variation (AMOVA) by watershed should support this pattern.

2. Regional Clusters:

AMOVAs on population clusters (cells) within watersheds should indicate an additional level of structuring within the watershed. Alternatively, clustering of haplotypes across watershed boundaries would suggest that contemporary watershed boundaries have not wholly restricted gene flow.

3. Glacial Refugia:

Population structure reflects the patterns of glaciation (advances and retreats), in concert with likely refugia (remixing or divergence zones). The genetic structure for older populations (i.e., southern) should be both

deeper and representative of multiple radiations, while younger populations (i.e., northern) should be relatively shallow with both unique and fewer haplotypes representing relatively recent dispersal events.

Each of these models will be examined to determine the best fit to the genetic data. The objective is to articulate a phylogeographic model for *M. lustrica* from which issues of conservation can be rationally addressed.

CHAPTER 2

LITERATURE REVIEW

Ecology

M. lustrica is a poorly understood aquatic snail, relatively rare throughout its range and listed in the State of Massachusetts as Endangered. Among the largest group of aquatic mollusks, *M. lustrica* is an operculate, prosobranch snail, part of the freshwater subfamily Nymphophilinae (Hydrobiidae), and containing 159 species in ten genera (Hershler et al 2003). It is the northern-most cold-temperate species of its genus, with other members of the genus occurring along the southern edge of its range (Fig. 2) (Thompson 1977). It is concentrated around the Great Lakes in Michigan, with fewer populations in the east, and occurs at relatively low frequencies in New York and Massachusetts (Fig. 1) (Berry 1943; Harman and Berg 1971; Thompson 1977; Jokinen 1992) and no populations in the rest of New England. It is widespread in Michigan, where it is possibly the densest, but there it is still only half as common as *Amnicola limosa* (Berry 1943). *M. lustrica* was listed as endangered in the State of Massachusetts in 1986 under the Massachusetts Endangered Species Act, but it is not listed at the federal level.

Our understanding of the ecology of *M. lustrica* is limited to a handful of state and regional snail surveys and taxonomies (Baker 1928; Berry 1943; Harman and Berg 1971; Strayer 1987; Jokinen 1992), as well as a few reports from Stockbridge Bowl (Ludlam et al 1973; Fugro 1996; ENSR 1998; Coote and Roeder 1999; Roeder and Coote 2000; McLain 2003). Most of these reports are limited, focusing on distribution and basic ecological information, such as whether the populations are found in ponds, streams or

lakes, and whether individuals are found on rocks or plants (Burch 1988 & 1989). The majority of work on the genus has been conducted by Dr. Fred Thompson focusing primarily on the southern species, and addressing their highly disjunct ranges throughout the southeastern United States (Thompson 1977; Hershler and Thompson 1987; Thompson and Hershler 2002; Hershler et al 2003).

Many malacologists assert that avian dispersal of aquatic snails is significant for the establishment and maintenance of populations (Jokinen 1983; Boag 1986). Such a dispersal process depends on eggs and juvenile snails adhering to wading birds (Boag 1986). The biology of *M. lustrica* does not lend itself to such a dispersal process, nor does its isolated distribution in western Massachusetts, where lakes, rivers, ponds, and waterfowl are all abundant, support such a process. Others have suggested that the distribution of the Nymphophilinae family indicate limited dispersal ability, often being restricted to specific drainages (Hershler et al 2003 and 2008).

M. lustrica is reported as typically found in association with *A. limosa*, with which it shares many characteristics, including size, number of whorls, color, and habitat (Berry 1943; Jokinen 1992). *A. limosa* is both ubiquitous and proliferate, occurring throughout much of North America, making the association of the two snails more coincidental than biologically meaningful. *M. lustrica* is dioecious and lays single eggs similar to *A. limosa*, but eggs of *M. lustrica* are round and lack a laminated crest (Berry 1943; Kesler 1980; Smith 1995). While *A. limosa* are annuals that deposit eggs from May to June, the life cycle of *M. lustrica* is uncertain (Jokinen 1992). However, the pattern of juvenile *M. lustrica* observed in two surveys in 1999 and 2000 (Coote and Roeder 1999, Roeder and Coote 2000) is consistent with an annual life cycle similar to that of *A.*

limosa; typically large adults are found in the early spring with a shift in numbers towards juveniles in the late summer and early fall.

Researchers have suggested that the chemistry of water bodies, especially regarding calcium availability and pH, plays an important role in the presence or absence of snails. Some research supports a correlation between water chemistry and the distribution of snails (MacNamara and Harman 1975, Jokinen 1983, 1992; Jokinen and BS-NYSM 1987; Økland 1990; Lewis and Magnuson 2000) but other research has found that chemistry plays a minor or secondary role (Harman and Berg 1971; Lodge et al 1987; Lewis and Magnuson 2000). There is, however, some consensus that calcium plays an important role as a limiting factor at the extreme; e.g., when calcium concentrations are < 5 ppm, most snails are unable to become established, presumably due to an inability to form sufficient shell structure (Lodge et al 1987; Jokinen 1992). In the case of *M. lustrica*, snails have been found in a wide variety of habitats including streams, rivers, and lakes (Berry 1943). Populations of *M. lustrica* have been found in medium to hard-water lakes and in freshwater marshes, ponds, and rivers, and are associated with submerged aquatic vegetation (SAV) including *Vallisneria*, *Potamogeton*, and *Chara* sp. and the invasive plant species *Myriophyllum spicatum* (Berry 1943; Harman and Berg 1971; Jokinen 1992; ENSR 1998; Roeder and Coote 2000; Massachusetts Natural Heritage and Endangered Species Fact Sheet 2010). I have also found *M. lustrica* on open substrate including mud, sand, and cobble.

It has been reported that the major habitat association for *M. lustrica* is with *Chara* sp. (Berry 1943, Ludlam et al 1973; Jokinen 1992; Smith 1995; Fugro 1996; ENSR 1998), but recent research suggests that its ecology is more complex, and that the

Chara-snail association is not with each other but within the general association of glacial lakes. Specifically, there are multiple *Chara* species, and within Stockbridge Bowl at least, the snail is found as frequently in association with other SAV species (notably *M. spicatum*) as in *Chara* beds (Coote and Roeder 1999; Roeder and Coote 2000; McLain 2003).

The vast majority of *M. lustrica* in Stockbridge Bowl occur in less than 2 m of water but are found in depths to 4 m (Ludlum et al 1973; Coote and Roeder 1999). Other studies have found the snail to be most abundant in 4-8 m of water in larger lakes (e.g., Lake Michigan), possibly due to wave action (Jokinen 1992). Some have suggested that *M. lustrica* may be migratory, moving from shallow waters in the summer to deep waters in the winter to avoid freezing (Jokinen 1983). Migration has been shown to occur with *A. limosa* (Horst and Costa 1975), while studies on other aquatic snails have not documented such movements (Wall 1977). The case for migration in *A. limosa* is based on the presence/absence of snails or eggs at various depths (Horst and Costa 1975; Kessler 1980), not on mark and recapture studies. My research on the movement of *M. lustrica* is inconclusive, (Coote and Roeder 1999; Coote and Schmidt 2005, unpublished) but suggests that movement to deep waters during the winter months is not universal, with individuals being found in less than 1 meter of water well into November and December. In one study the shoreline was sampled through the month of January, collecting samples in less than 2 meter of water through a foot or more of ice. Individuals were present throughout the study in less than 1 meter of water, with no indication of movement (i.e., changing densities) to deeper water. The speculation that *M. lustrica* is

migratory is based largely on its suspected association with *A. limosa* and does not appear to be a significant biological trait.

The extent of our ecological knowledge of *M. lustrica* is relatively limited, not due to a lack of studies or effort, but due to its generalist nature, low incidence and fecundity, and minute size (3-6 mm). Given the general decline of snail species in the United States and worldwide, as well as when one considers the loss of habitat associated with that decline (Burch 1989; Lydeard et al 2004; IUCN 2004), the chance of extinction of *M. lustrica* is real. Considering the importance of snails to the ecology of lakes (Kabat and Hershler 1993), the contribution of rare invertebrates to species richness in aquatic systems (Cao et al 1998), and our relative lack of gastropod knowledge, it seems logical to develop a greater understanding of this snail in the context of long-term environmental change and climate change in particular.

Taxonomy and Morphology

The taxonomy of Nymphophilinae based on morphological characters is messy, with species exhibiting phenotypic plasticity and many convergent overlapping character states (Wilke et al 2001 & 2002; Dillon & Frankis 2004). The shells are cryptic, and the extent to which the internal anatomy is helpful in identification is debatable (Wilke et al 2001 & 2002; Hershler et al 2003 and 2008). This is evidenced by the wide array of names given by various authors to the same species, the inclusion of the same species in different genera during the past century and a half, the increasing number of species “discovered” in the past 50 years, and the several taxonomic revisions, first based on

morphology, and now on genetic analysis (Harman and Berg 1971; Thompson 1977; Hershler and Thompson 1987; Hershler 1994; Hershler et al 2003).

With regards to *Marstonia*, prior to 1977 there were only three recognized species: *olivacea*, *agarhecta*, and *lustrica* (Thompson 1977). Thompson described five new species and also placed five names in synonymy with *M. lustrica*: *Amnicola oneida*, *A. winkleyi*, *A. perlustrica*, *M. decepta*, and *M. gelida* (ibid). Since that time 6 more species have been described or reassigned from *Pyrgulopsis*, with some new ones currently being described (Coote current; Hershler pers. comm.; Perez et al 2005; Hershler et al 2003; Thompson and Hershler 2002).

In the case of Hydrobiids, morphological characters are useful but are limited in many cases to analysis at the family level or higher, with many of the characters being plastic or cryptic, and lacking the synapomorphies necessary to confidently delineate among species (Wilke et al 2001). Wilke et al (2002) tested both qualitative (from Hershler and Ponder 1998) and quantitative (from Davis et al 1992) morphological character states typically used for hydrobiid snails by dissecting 75 specimens from seven populations and four taxa of *Hydrobia* sp. None of the qualitative character states were useful for delineating among the taxa and quantitative states measured were not useful for differentiating species. However, by employing a discriminate analysis using mtDNA lineages as groupings, shell characteristics performed better than soft body parts and the overall performance of the models was highly significant. Other researchers have also found that in the case of cryptic gastropods, using traditional methods are helpful but are not definitive (Strothard et al 1996; Hershler et al 2007; Hershler et al 2008). For these

reasons, genetic analysis has become a critical component of the development of Hydrobiid taxonomies.

Marstonia spp are cryptic, difficult to tell from other hydrobiids based on shell structure alone, and are difficult to differentiate based on soft body parts. While Hershler and Ponder (1998) and Davis et al (1992) have identified several character states that can be used for morphological analysis among hydrobiids (including *Marstonia*) and rissoodideans, respectively, Wilke et al (2002) showed that only five quantitative characters are useful for delineating species in conjunction with genetic lineages among the closely related *Hydrobia*, that the family as a whole remains cryptic, and insisted that genetic analysis is a prerequisite for taxonomic work on cryptic snails.

Biogeography across the Range

The current southern edge of the range of *M. lustrica* follows the front of maximum expansion of the Laurentide ice sheet with its current full range extending from Indiana and Ohio northwest to the upper Mississippi River drainage in Minnesota, through the Great Lakes states, and east to southern Ontario and western Massachusetts (Ludlam et al 1973; Thompson 1977; Strayer 1987; Jokinen 1992; Hershler 1994) (Fig. 1).

It is suspected that the origin of *M. lustrica* is the Tennessee River refugium (now extinct there), having radiated out from the Ohio Valley with the retreat of the Laurentide glacier (Thompson 1977, Strayer pers. comm.), accessing central New York from the Great Lakes, and accessing the Hudson Valley via the Mohawk Valley and/or the Erie Canal (Schmidt 1986: Strayer 1987). The map in Figure 1 is from Thompson (1977), and

since much of the work on *Marstonia* has taken place in the past 20 years, with several newly described species, the map should be viewed with caution.

In 2002 Thompson and Hershler resurrected *Marstonia* to generic status and assigned all *Pyrgulopsis* east of the Mississippi to it. Using genetic markers, Hershler et al (2003) revised the phylogeny of the Nymphophilinae establishing an eastern and western fauna, reinforcing the earlier division of Thompson and Hershler (2002). A notable exception of this division is *Floridobia*, which is nearly identical to *M. lustrica* in shell morphology, and may have confounded early reports of *M. lustrica* along the northeastern seaboard, particularly in the Hudson River and north to southern Maine. Thompson and Hershler speculated that the presence of *Floridobia* on the eastern seaboard represents an invasion along the coast from the Gulf of Mexico, and link this invasion to Laurentide flooding events (ibid).

There is a clear divide between the eastern seaboard and the main range of *M. lustrica*, delineated by the Allegheny and Appalachian ranges. This delineation is apparent as populations of *M. lustrica* follow along the edge of Lake Erie, clearly dipping down into Ohio but do not cross over these ranges until eastern New York. To the west, the range of *M. lustrica* does not appear to be confined by mountains, but does border along the upper Mississippi in Iowa and Minnesota.

Other species of *Marstonia* are spread thinly throughout the Midwest and southeastern United States, but are generally confined to east of the Mississippi (Fig. 2 and Table 1). There are a few records of two species, *M. letsoni* and *M. scaliformis*, sharing the southern range of *M. lustrica* (Burch 1982), along the southern Michigan and

northern Ohio border, but there are no published mtDNA sequences for these species. The remaining species are also rare, approximately half of which are sequenced, occurring in only a few localities from eastern Georgia west to the states along the Mississippi, with one disjunct population in Texas. Currently the genus is divided into three general groupings with *M. lustrica*, *M. letsoni*, in the north/northeast, *M. scaliformis*, *M. pachyta*,

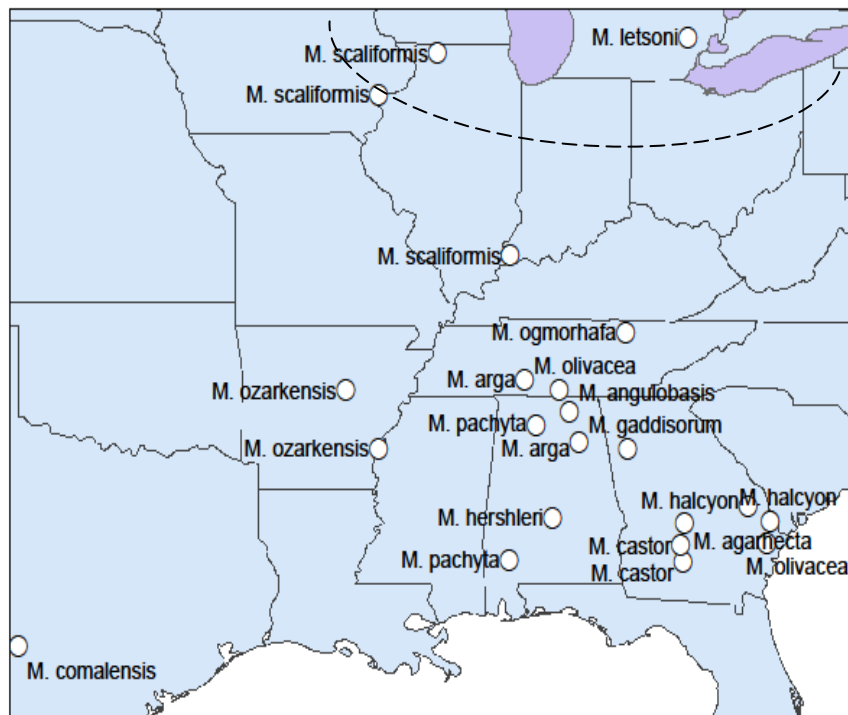


Figure 2 – Map of other *Marstonia* species. Locations of current records for the species are approximate. Dashed line indicates approximate southern border of *M. lustrica*.

M. arga, *M. ogmorhapha*, *M. hershleri*, and *M. olivacea* in the midwest (Upper Mississippi, Ohio, and Tennessee valleys), and *M. castor*, *M. halcyon*, *M. agarhecta*, *M. gaddisorum*, and *M. angulobasis* in the southeast (Table 1). *M. ozarkensis* and *M. comalensis* appear to be disjunct populations west of the Mississippi.

Biogeography of the Eastern Populations

While understanding the impact of glaciation on the general distribution of *Marstonia* is a significant part of this dissertation, equally important is understanding its distribution across the eastern range. As recently as 1987, the eastern range of *M. lustrica* was not considered to extend beyond the Hudson River (Strayer 1987)³. The range of *M. lustrica* now includes two confirmed locations in western Massachusetts: glacial lakes Stockbridge Bowl, Stockbridge, MA and Laurel Lake, Lenox/Lee, MA a few miles east of Stockbridge Bowl (Roeder and Coote 2000; Coote 2005, unpublished). Several other lakes surveyed in western MA and eastern NY yielded no evidence of the snail, despite the appropriateness of the habitat (i.e., glacial lakes with *A. limosa* and *Chara*).

The snail is widely scattered across central New York (Harman and Berg 1971; Jokinen 1992) with a few populations in the Hudson River Basin (Strayer 1987; Jokinen 1992). The snails found in the Hudson Basin are restricted to the Hudson River and a few locations directly adjacent to the river. Previous work on *M. lustrica* in New York is restricted to the snail taxonomies and surveys covering the state (Harman and Berg 1971; Strayer 1987; Jokinen 1992), which include basic habitat associations and some water chemistry data. There are several works from the 1800s and early 1900s in New York of similar nature but none of these discuss the ecology or biogeography of this snail in any detail (Lewis 1856, 1860, 1868, 1874; Pilsbry 1890; Baker 1928). Harman and Berg

³ There has been considerable confusion concerning the nomenclature of *M. lustrica*, having been renamed several times in the past 50 years. While most taxonomies have placed its easternmost range, currently and historically, in eastern NY, we now know that it was present in western Massachusetts. It also appears that there was once a population in the Housatonic River in Connecticut under the name *M. decepta* which was declared extinct in 1981 (Baker 1928; Jokinen and Pondick 1981). The population in Stockbridge Bowl was not formally recognized until the late 1970s and did not make it into any formal taxonomies until the 1980s.

Table 1 – List of other *Marstonia* species.

Species	Region	State
lustrica	north	MN, MI, MA, NY, Canada, OH
letsoni	north	MI
scaliformis	midwest	IL, IN, IO, MO
pachyta	midwest	AL
arga	midwest	AL, TN
ogmorhafa	midwest	TN
hershleri	midwest	AL
angulobasis	midwest	AL
gaddisorum	midwest	GA
agarhecta	southeast	GA
castor	southeast	GA
halcyon	southeast	GA
olivacea	southeast	AL, GA
comalensis	disjunct	TX
ozarkensis	disjunct	AR

(1971) and Jokinen (1992) did substantial fieldwork on gastropods, collectively covering the entire state. Strayer (1987) reviewed the records of mollusks of the Hudson Basin and examined thousands of museum specimens, and confirmed several records of this snail in the Hudson Basin based on shell morphology, which Jokinen's fieldwork in 1992 corroborated. Two populations have been reported east of the Hudson River including the Saw Kill River (Dutchess Co.) and the Williams Bridge (Westchester Co.) (Strayer 1987) but the later location was not confirmed by Jokinen (1992). It is important to note that without genetic analysis, the tidal or brackish populations identified as *M. lustrica* in the Hudson River prior to 2008, in the absence of the knowledge of the presence of *F. winkleyi*, are also suspect.

The distribution of *M. lustrica* in eastern New York and western Massachusetts is somewhat puzzling. While the distribution of the snail in western and central New York is consistent with the pro-glacial (during) and post glacial (after) hypothesis, with the species thinly but widely scattered across the region, the presence of the snail in eastern New York is not easily correlated with the retreating ice. The main issue is that the presence of the snail in the Mohawk and Hudson Valleys is limited to a few locations along the Mohawk and Hudson Rivers and is not found in lakes far beyond the main rivers as would be expected if its presence was part of the expansion and contraction of pro-glacial lakes. Confounding the issue is the Erie Canal. Strayer (1987) has suggested that the distribution pattern may be the result of the Erie Canal and that the presence of the snail in the Hudson Valley is relatively recent. Complicating matters is the recent discovery of *Floridobia winkleyi* in the upper reaches (non-brackish) of the Hudson River, throwing into question earlier identifications (Coote and Strayer 2008; Davis and Mazurkiewicz 1985).

The eastern distribution of the snail becomes more complicated with the addition of the Massachusetts populations, which are approximately 300 m above the Hudson River, and separated from the Hudson drainage by the Taconic range. There is some evidence for the dispersal of aquatic fauna from the Hudson Valley to the Housatonic Valley via pro-glacial Lake Albany and Lake Bascom, through the Hoosic River valley (Fig. 3) (Bierman and Dethier, 1986), but such a dispersal route would be problematic given the relatively immobile nature of the snail, requiring active upstream movement over long distances and relatively short time frames. It also raises questions about distribution of *M. lustrica* in southwestern Massachusetts, as opposed to the northwest

part of the state, as it is not reconcilable with the glacial history of western Massachusetts. In short, in order to reach the Housatonic drainage, Lake Albany and Lake Bascom would have acted only as stepping stones for *M. lustrica*, an unlikely scenario given its biology.

Specifically, glacial Lake Bascom covered the northwest corner of Massachusetts south to Berkshire, MA (Fig. 3) (Bierman and Dethier 1986). It did not include Stockbridge Bowl or Laurel Lake in the south (Glacial Lake Housatonic), but did flow south for a period of time. However, when Lake Bascom flowed south over the spillway at Berkshire, MA (317 m) into the Housatonic Valley, Lake Bascom was not connected to Lake Albany. Not until the ice retreated further northwest in New York did a second Lake Bascom spillway open at Potter Hill, NY (273 m), which connected Lake Bascom with Lake Albany via the Hoosic drainage. There is evidence that Lake Albany eventually controlled Lake Bascom levels, but not until levels drained to 213 m (ibid), at which point, Lake Bascom was no longer connected to the Housatonic drainage.

Stockbridge Bowl and Laurel Lake are several miles south of Berkshire, MA and thus beyond the southern extent of Lake Bascom. During the time that Lake Bascom flowed south it was prior to a connection to Lake Albany. By the time Lake Albany and Lake Bascom were connected, Lake Bascom was no longer connected to the Housatonic drainage or Stockbridge Bowl or Laurel Lake. At this point it seems unlikely that *M. lustrica* would have reached these lakes via the Hoosic drainage as part of Lake Albany. It has also been recognized that the mollusk species found in the Hudson Valley are significantly different than those found in New England, which has been explained as a result of the Taconic and Green mountain barriers (Smith 1983; Jokinen and BS-NYSM

1987), which undermines the case for a connection for *M. lustrica* between the two watersheds. If we accept the historical record of *M. lustrica* in the Housatonic drainage in Connecticut (Linsley 1845, in Smith and Prime 1870; also in Jokinen and Pondick 1981), it seems likely that the presence of the snail in western Massachusetts is not connected to the Erie Canal or otherwise introduced from the Hudson Valley in recent times, but is somehow associated with a pro-glacial lake system other than Lake Albany, with populations surviving glaciation in southern refugia, possibly Lake Connecticut.

Another possible explanation is a pro-glacial lake with an extent not yet discovered, which covered eastern New York and was connected to the Berkshire Valley over the Taconic and Berkshire ranges. There is support for a general connection along the front of the retreating glacier around 12,000-13,000 yr BP, with isochrones extending along the border of Pennsylvania and New York through the northwest corner of Connecticut (Bryson et al 1969). One such connections may have existed on the border of New York and Massachusetts at West Stockbridge, MA (pers. obs.), and another may have existed at Ten Mile River on the border of New York and northwestern Connecticut, the latter of which connects *Exoglossum maxillingua* (cutlips minnow) populations in the Hudson and Housatonic watersheds (Schmidt pers comm.).

If *M. lustrica* in Massachusetts are not a result of pro or post-glacial processes via a connection between Lake Albany and Lake Bascom, then one explanation for the distribution of the snail into Massachusetts and the Housatonic Valley could be an advancing population from the south.

In summary, the late Wisconsin period consisted of numerous glacial advances and retreats, creating large and small pro-glacial lakes across the front of the glacier,

across the entire contemporary range of *M. lustrica*. These bodies of water experienced both brief (<100 years) and long (>1000 years) life cycles, experienced stable and catastrophic flow regimes, and periodically flowed to the north, south, east, and west. Despite what appears as a chaotic environment of constant change, there are general patterns in lake formation and flow regimes which can be broken down by region and time period.

Glaciation

It is well known that global climate cycles and glacier advance and retreat have played a significant role in the distribution and evolution of a number of plants and animals (Schmidt 1986; Hewitt 1995, 2000 and 2004; Yang et al 2001; Rowe et al 2004; Emerson and Hewitt 2005; Curry 2006). The specifics of how these forces impacted individual species are not well understood, yet many species under protection and subject to conservation efforts are likely to be exhibiting relic behaviors and distributions that can only be understood in the context of glaciation and evolutionary history (Rowe et al 2004).

The Laurentide glacier was one of the largest glacier systems, episodically covering most of North America south to 40° latitude from 0.1-2.5 mya (Larson and Schaetzl 2001; Table 2 and Fig. 3). The major events include the Wisconsinan from 10-80 kya, the Sangamon interglacial 80-130 kya, the Illinoian 130-310 kya, and the pre-Illinoian periods from 0.3-2.5 mya (Fullerton et al 2004; National Atlas 2010). The

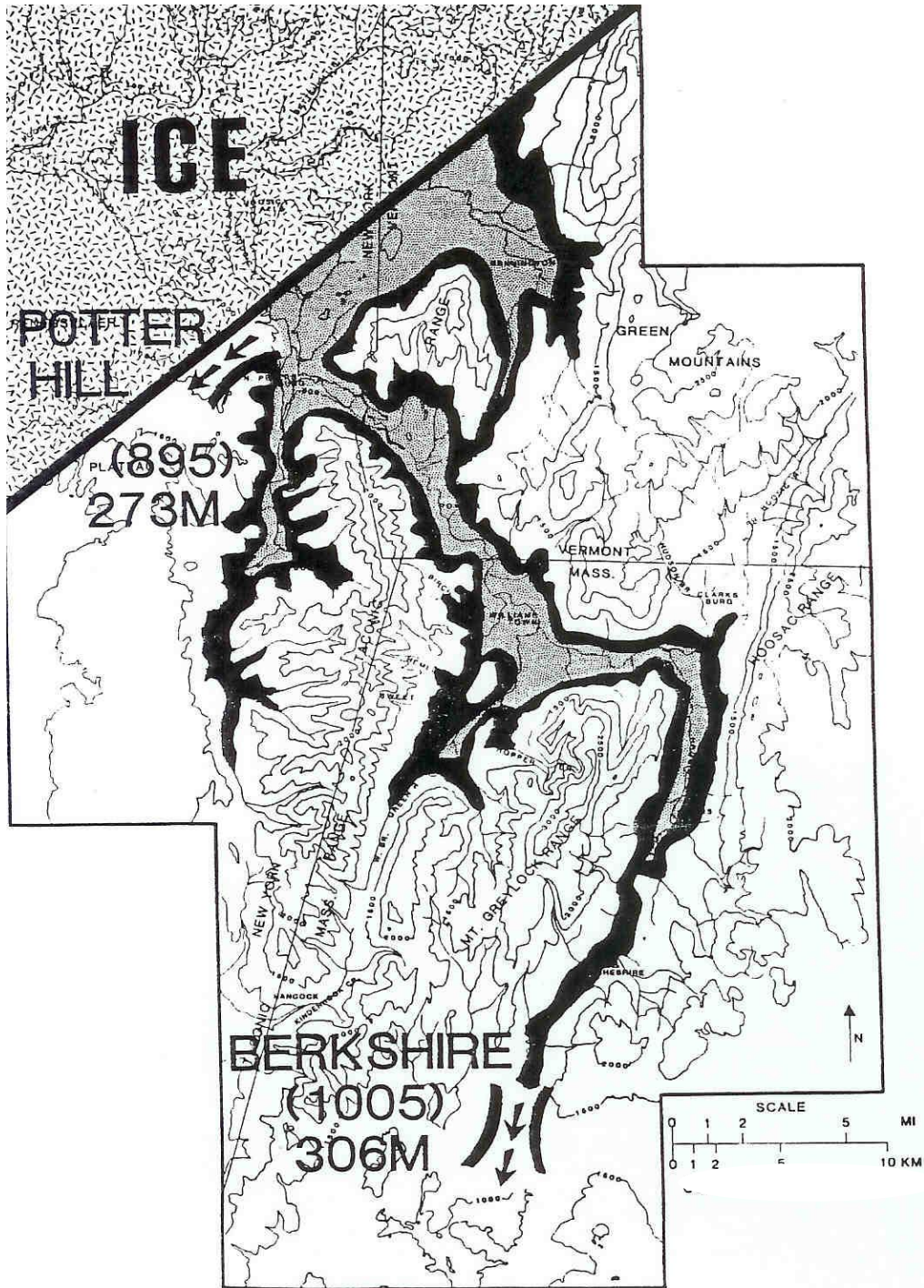


Figure 3 – Map of glacial Lake Bascom. An idealized representation of Lake Bascom and the retreat of the Wisconsin Ice Sheet from northwestern Massachusetts and eastern New York. Arrows indicate spillways. Flow was at first south, then with retreating ice flowed west into Lake Albany (Bierman and Dethier 1986).

specific details of the periods prior to the Wisconsinan are few due to the obscuring effect of the Wisconsinan glacier.

The Wisconsinan had a maximum extent to about 41° latitude, dating to about 20 kya. The Illinoian and pre-Illinoian extended slightly further south to about 38° latitude, and covered a much greater area longitudinally (Fig. 4). These advances and retreats are significant because we know that any *Marstonia* present 0.3-2.5 mya would have been pushed back to the Tennessee Valley at least twice by the Illinoian and pre-Illinoian, and then again to the Ohio and upper Mississippi several times during the Wisconsinan. Given the current range of *M. lustrica*, clearly glaciation has played a significant role in its evolutionary history.

The Wisconsinan itself was made up of three main time periods: the Tahoe is estimated to have been at its maximum extent at approximately 70 kya, followed by the Tenaya estimated to have existed from 30-50 kya, and finally and most recently, the Tioga with its maximum extent estimated to have occurred between 23-18 kya (Eschman 1985; Hewitt 2004). Each of these major periods can be broken down into additional substages and interglacials (Table 2).

The Wisconsinan was stable at or near its maximum expansion on four occasions (Sugden 1977; Fullerton 2004) forcing *M. lustrica* back into the upper Mississippi and Ohio valleys, or possibly along the front of the glacier on the Atlantic seaboard. Following each advance was a significant retreat, or interglacial period, each developing large pro-glacial lakes and water systems across the front of the glacier (Appendix H). After the initial advance of the Tahoe 50-80 kya, there is evidence of a substantial retreat

known as the St. Pierre interglacial (Barnett 1992). Following this retreat is the Guildwood substage dating to about 60 kya. A significant retreat then follows known as the Port Talbot interglacial, uncovering most of Michigan, Ohio, and Indiana, as well as the St. Lawrence drainage, dating from 40-55 kya (Eschman 1985, Barnett 1992). This retreat marks the most recent major potential re-expansion period for *M. lustrica*. The ice then re-advanced covering most of its previous range, culminating around 23 kya in the Cherrytree substage. This is followed by yet another retreat (Plum Point interglacial), then another significant advance, the Nissouri substage, marking the final maximum of the Laurentide ice 20 kya (Mayewski et al 1981; Eschman 1985; Barnett 1992). At this

Table 2. Glacial events and time frames.

Period	Substage	Time
pre-Illinoian		0.3-2.5 mya
Illinoian		130-310 kya
Sangamon	(interglacial)	80-125 kya
Wisconsinan		10-80 kya
	Tahoe	50-80 kya
	Nicolet substage	80 kya
	St. Pierre interglacial	75 kya
	Guildwood substage	60 kya
	Tenya	22-50 kya
	Port Talbot interglacial	40-50 kya
	Cherry Tree substage	35-40 kya
	Plum Point interglacial	24 kya
	Tioga	18-22 kya
	Nissouri substage	20 kya
	Erie interstadial	15-18 kya
	Port Bruce substage	14 kya
	Mackinaw interglacial	13.5 kya
	Port Huron substage	13 kya
	Two Creek interglacial	12 kya
	Greatlaken substage	11.9 kya
	Nipissing phase	5.5-11.9 kya
	Post-Nipissing phase	present

(Barnett 1992; Eschman 1985; Mayewski et al 1981; Benn and Evans 1998)

time most of Ohio, Indiana, and Illinois were covered by ice, as were all of New York, the northern edge of Pennsylvania and New Jersey, and all of New England. All of Wisconsin, except for the driftless region, and Minnesota were covered as well. Most, if not all, land-based and aquatic animals in ranges covered by the ice had been extirpated, reduced to refugia, or forced to migrate south in front of the advancing ice sheet. After the Nissouri substage, multiple shorter coolings and warmings occurred between 21 kya and 10 kya (Eschman 1985; Holman 1992), each in its turn promoting and then trimming

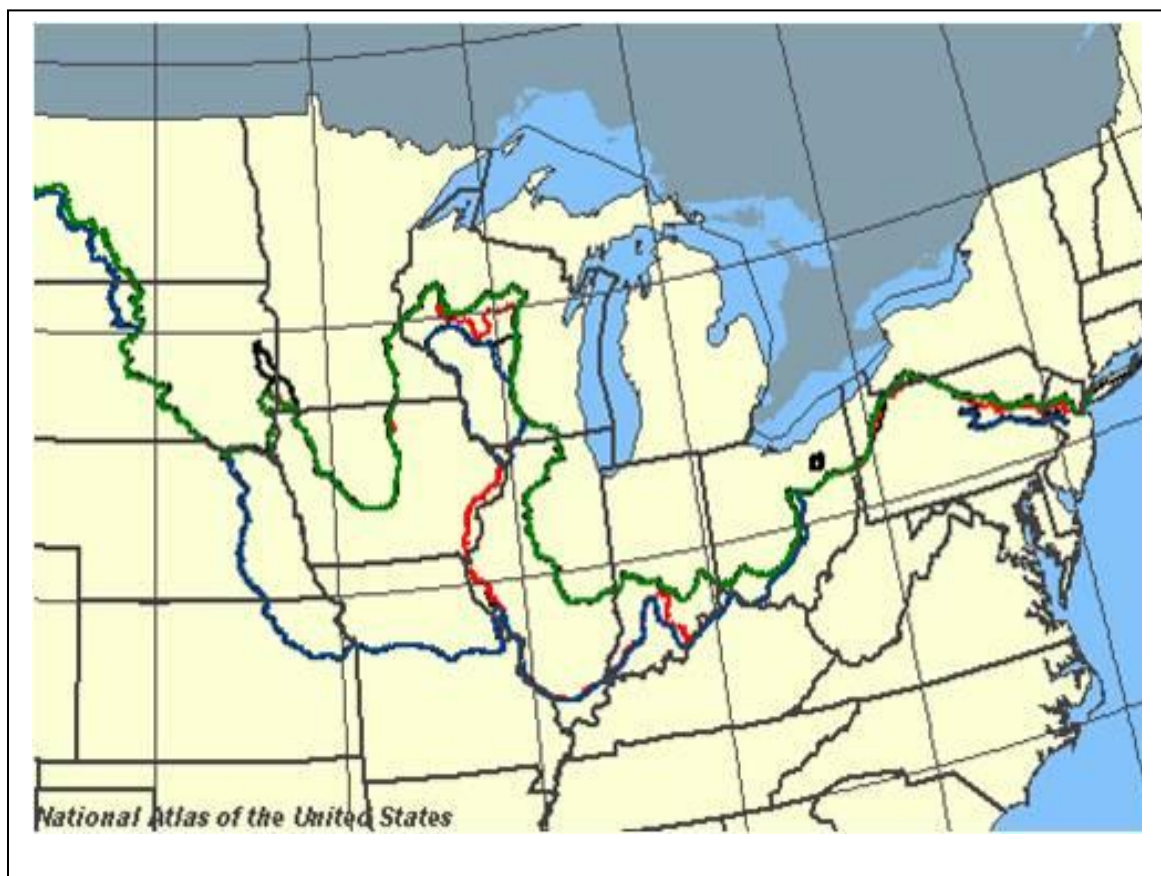


Fig. 4 – Map of major glaciations. Map shows the maximum extent of the Laurentide Glacier and successive series of advances. The blue and red lines show the maximum extent of the older Illinoian and pre-Illinoian glaciations, respectively, 0.13-2.5 mya. The green line indicates maximum extent of the Wisconsinan 20 kya.

the radiation of numerous species. Early forms of the Midwest lake systems may have developed as early as 21 kya but lakes in the Northeast were certainly forming, including Lake Hitchcock, CT and Lake Connecticut (Long Island sound) around 18 kya (Rittenour 1997; Lewis 1997). There is no evidence of *M. lustrica* ever existed further east than the Housatonic Valley but Lake Connecticut or other eastern seaboard lake may have acted as a source for the snail in the Housatonic Valley.

At approximately 16 kya significant warming occurred and the ice retreated again (Erie interglacial), exposing land roughly to the border of Canada and creating water bodies from across the Great Lakes region east to the Atlantic seaboard. The Erie interstadial (minor retreat) was then followed by the Port Bruce substage 14 kya, with ice re-advancing to the northern edges of Illinois, Indiana, Ohio, once again covering the Great Lakes, but leaving the southern range of *M. lustrica* unglaciated (Eschman 1985). The final retreat of the ice from the current range of *M. lustrica* at approximately 10 – 12 kya marks the onset of the current interglacial (Eschman 1985; Holman 1992).

The end of the Port Bruce substage at 15 kya marks the beginning of sustained ice retreat and substantial development of pro-glacial lakes, including Lake Maumee (Lake Erie and southern edge of Lake Huron) and Lake Chicago (Lake Michigan) (Table 3). Hypothesized glacial maps for each of the lake systems described here, with flow regimes, can be found in appendix G. At approximately 14,100 yr BP, a re-advance forced Lake Maumee to flow west into Lake Chicago via the Saginaw Basin with an outflow southwest into the Mississippi drainage. This period was followed by a significant retreat known as the Mackinaw Interstadial, lowering the water level in the Lake Michigan, Erie, and Huron basins, possibly lower than contemporary levels, with

evidence that the flow out of Lake Maumee was reversed, flowing east into the Mohawk Valley (Eschman 1985).

An important re-advance was the Port Huron substage at 13 kya, covering all but the southern end of Lake Chicago, all of current day Lake Huron and forming Lake Saginaw, covering current day Lake Ontario, and forming Lake Whittlesey (Lake Erie) (Eschman 1985; Schaetzl et al 2000; Siegert 2001). At this time Lake Whittlesey flowed into Lake Saginaw, which in turn flowed into Lake Chicago. Approximately 12 kya the ice then retreated, exposing these basins once again, forming early Lake Algonquin, by

Table 3 – Major glacial lakes, time frames, and flow regimes.

Name	Contemporary Reference	Time Period (kya)	Flow Regime
Lake Connecticut	Long Island Sound	18-20	east ?
Lake Albany	Hudson River	8-14	south
Lake Bascom	northwest Massachusetts	12-14	south then north
Lake Maumee	Lake Erie	12-14	west
Lake Chicago	Lake Michigan	11-12	west
Lake Warren	Lake Erie	11-12	west to Lake Chicago and east
Lake Algonquin	Lake Michigan	10-11	east to Lake Iroquois and west
Lake Iroquois (Frontenac)	Lake Ontario	10-11	east to Hudson, and north east
Lake Duluth (Beaver Bay)	Lake Superior	9-10	west to upper Mississippi
Main Algonquin	Lakes Michigan and Huron	9-10	south to Lake Erie
Lake Erie	" "	9-10	east to Lake Ontario
Lake Ontario	" "	9-10	northeast to St. Lawrence
Lake Superior	" "	8-9	west and east
Lake Algonquin	Lake Michigan	8-9	southwest to Mississippi and east
Lake Erie & Ontario	" "	8-9	northeast to St. Lawrence
St. Lawrence River	" "	8-9	south to Hudson?
Lake Superior	" "	4-6	east
Lake Nipissing	Lake Michigan and Huron	4-6	east to Erie and south to Mississippi
Lake Ontario	" "	4-6	to St. Lawrence and Hudson?

connecting Lake Chicago with Lake Saginaw. At approximately 11.8 kya the ice re-advanced (the Greatlakean substage) reforming Lake Chicago out of early Lake Algonquin, covering the northern halves of current day Lake Michigan and Lake Huron (Eschman 1985; Holman 1992; Schaetzl et al 2000). During this re-advance, water levels eventually rose in these basins, extending along and connecting Lake Chicago and early Lake Algonquin at the glacial margins to form the massive main Lake Algonquin (ibid). Lake Algonquin flowed due east across the front of the glacier to the St. Lawrence drainage, and at this time was not connected to early Lake Erie. Early Lake Erie flowed east into Lake Ontario (Schaetzl et al 2000), or possibly drained through central New York to the Hudson River via the Rome outlet (Mayewski et al 1981). There is additional evidence that during this period, specifically between 11-13 kya, there may have been a much larger body (or bodies) of water covering the region at an elevation higher than Lake Algonquin, possibly connecting all the major basins across the front of the glacier (Schaetzl et al 2000). Evidence for marine intrusion at this time includes Saint Lawrence marine deposits dated to 12-13 kya as far west as the Ottawa River valley, and marine submergence for eastern Maine for the same time period (Mayewski et al 1981). Any water bodies in those zones that were previously fresh would have become saline, further trimming fresh water aquatic species on the eastern seaboard.

Around 11-12 kya Lake Bascom, Lake Connecticut, and Lake Hitchcock drained. Around 10 kya the Champlain Sea turns into Lake Lampsilis due to eustatic rebound, Lake Duluth (Lake Superior) becomes established flowing into the upper Mississippi, and Lake Algonquin expands significantly with two outlets one into Lake Erie and one out of Lake Michigan flowing into the Upper Mississippi. Lake Ontario is also established by

this time and is likely flowing east into the Mohawk and Hudson basins through the Rome outlet, and later into the St. Lawrence. Lake Albany has now drained. By 8 kya Lake Superior is established and flowing either east into Lake Algonquin or continues to flow out west into the upper Mississippi. By 6 kya Lake Superior is no longer flowing west but is now flowing into Lake Nipissing (formerly Lake Algonquin). Contemporary flow patterns and Lake Huron become established in the past 6 kya.

With the final retreat of ice from the area with the end of the Greatlakean substage, most of the lakes drained north into the St Lawrence Seaway occupying the Ottawa River basin (Eschman 1985). At approximately 11 kya, ice advanced across the eastern half of Lake Superior basin forming Lake Duluth, which flowed west out the St. Clair basin. At this time Lake Algonquin encompassed Lake Michigan and Lake Huron, flowing north east along the front of the glacier through the Ottawa River basin, and Lake Erie flowed into Lake Ontario (Schaetzl et al 2000). Final deglaciation around 9 kya resulted in significantly lower water levels, forming the beginnings of the contemporary Great Lakes. At approximately 4.5 kya, rebound of North Bay Ontario resulted in the impoundment of the Lake Huron and Lake Michigan basins, forming the massive Nipissing Great Lakes, with the current Great Lakes being established approximately 1-2 kya, which also changed the flow of Lake Superior from west to east.

Throughout the Wisconsin period, rivers and drainage basins were blocked or changed direction, and pro-glacial and post-glacial lakes and water bodies formed and dissipated (Hewitt 2000; Siegert 2001). During this time the depression of the land by the ice, which was significant (up to a mile or more), contributed to the pooling of water along the front of the entire glacier (Siegert 2001). These changes promoted or inhibited

expansion out of the Ohio Valley into the northern territories by freshwater obligate species (Hewitt 2000).

Unlike the generally accepted pattern of retreat in the Midwest, ice dynamics in New England are more complicated. It appears that instead of an active retreating ice front, retreat in New England resembled a combination of stagnant ice and an actively retreating ice front (Koteff and Pessl 1981). The implications to faunal distribution are significant; as such a structure of retreat would necessarily reduce the extent of connected water bodies across the front of the glacier, with stagnant ice possibly preventing the expansion of some species from advancing, with glacial lakes and potholes forming in situ and independent of the front. Early pro-glacial lakes in the northeast at the maximum front of the glacier and probably associated with the active retreat of ice include Lake Albany, Lake Bascom, Lake Housatonic, Lake Connecticut (Long Island Sound), and Lake Hitchcock in central Massachusetts. What is unknown is the extent of water bodies that would have presumably existed across the front of the glaciers during the height of advance, and their exact nature or duration.

Taxonomy and Genetics

Historically, the taxonomy of gastropods has been based on a combination of shell features, such as the number of whorls of the shell, the depth of the sutures, or whether or not it has an operculum. This basic method is generally considered inadequate for many taxa due to the plasticity (change in form due to environmental factors) of shell forms therefore other methods need to be used (Perez & Minton 2008). In the past century, soft body parts have played a central role in distinguishing among taxa, and

although useful (if not definitive), they can be less than helpful at the species level or below (Wilke et al 2002; Hershler et al 2008). One of the key problems with both approaches is the subjectivity involved, including the “lack of uniform(ity) in data sets, the subjectivity of an author, disagreement over character utility, and explicit or implied species-concept differences” (Perez & Minton 2008).

The advent of genetic analysis has been a boon in the delineation and identification of cryptic snails. Because *M. lustrica* is the most widespread of its genus, and given the lack of attention given the species as a whole, the use of genetic analysis provides an excellent opportunity to elucidate the phylogenetics of the species across its range and further our understanding of the genus’ relationship to the landscape. Faster and more objective than dissection, genetic analysis can provide answers to questions of relatedness, and can be used to infer phylogeography as well.

The use of phylogenetic analysis, including phylogenetic trees and haplotype networks, can infer the biogeographic history of organisms, as well as the landscape ecology of species at the macro level. Specifically, such information can be used to “reveal... historical barriers, geographic (e.g., peripheral) loci of differentiation, and patterns of gene flow” (Knowles et al 1999). In this particular case, the combination of methods used creates a model for understanding the phylogeography and landscape ecology of *M. lustrica*, with similar insight into the genus as a whole.

In addition to phylogenetic trees and haplotype networks, generalized time frames (molecular clocks) can be applied to these structures for the purpose of inferring associations with hypothesized events or to explore possible reasons for divergence. There are two ways to discuss molecular rate change: the first and early use was to use

the total branch length as divergence rate. Wilson and Sarich (1969 in Wilke et al 2009) used total branch length to suggest a generalized 2%/myr clock for mtDNA. Most contemporary authors use node depth as the substitution rate, which results in divergence rate divided by 2 = 1%/myr (Wilke et al 2009). Commonly reported rates for mtDNA marker cytochrome c oxidase I (COI) are 0.7-1.2% for corrected substitution rates, and 1.4-2.4% for uncorrected divergence rate/myr, and Wilke et al (2009) suggested a substitution rate of 1.18%-1.76%/myr (uncorrected) for mollusks. Wilke et al (2009) suggested that the COI fragment not be used to date phylogenetic events less than 200,000 years old. However, that is not to say that COI haplotypes derived in under 200,000 years are not informative regarding biogeographical structure of a species, and certainly can be used with caution (Hershler et al 2004a).

Using COI, Hershler et al (2003) clearly articulated the relationship of seven *Marstonia* (*M. lustrica*, *M. pachyta*, *M. hershleri*, *M. comalensis*, *M. halcyon*, *M. castor*, *M. agarhecta*) to other nymphophiline. They used *Phrantela marginatai* to root all trees, clearly showing the western fauna consisting of *Pyrgulopsis*, *Floridobia* (secondary invasion through the Gulf of Mexico), and *Nymphophilus*, while the eastern fauna consists of *Marstonia*, *Cincinnatia*, *Notogillia*, *Spilochlamys*, *Rhaphinema*, and *Stiobia*. Their analysis supports a loose grouping within *Marstonia*, placing *M. lustrica*, *M. pachyta*, *M. comalensis*, and *M. hershleri* together representing the midwest/northern species, while *M. halcyon*, *M. agarhecta*, and *M. castor* are grouped together representing the southeastern species. This builds upon Hershler's earlier morphological work where he delineated several *Marstonia* spp., and corroborated it, while giving insight to the position of some species not yet sequenced. Pairwise analysis demonstrates that the

specimens sequenced for *M. lustrica* differ from other species of *Marstonia* by 11-44 bp out of 550-650 bp. This level of differentiation is significant, is consistent with the geographic isolation of the different species, and is supported by morphological analysis.

Similar levels of structure and geographical variation have been found for other species. A study using COI on a widespread North American beetle *Ophraella* spp. indicated that there was shallow structuring in *O. communis* (1.04-3.6%) with 45 haplotypes (35 singletons) from 92 individuals, but showed no distance relationship. Its northern, more recently divergent sister species *O. bilineata*, occupying a similar range to that of *M. lustrica* resulted in 19 of 22 samples exhibiting unique haplotypes while being differentiated from the southern species by 1.97% (Knowles et al 1999). In this case, the lack of derived haplotypes being widespread, while the shared haplotypes were located on internal nodes of the trees, with a lack of a distance relationship, were used to infer that gene flow was not being driven by contemporary factors, suggesting historical processes. Taken together, the high number of singletons for both species indicates a history of episodic geographic expansion (glaciers), not of subdivision (vicariance) (ibid).

Like the beetle paper, a case on coastal estuarine amphipods (Kelley et al 2006) showed relatively shallow divergence with multiple singletons within and across populations, and a lack of a significant distance relationship. However, divergence across some samples was quite deep, indicating a new species in the southern range. Similarly, they found high diversity in northern, previously glaciated populations consisting of numerous unique haplotypes, which they explain by the likely existence of glacial refugia and subsequent remixing. Of particular interest here is the author's analysis of estuarine isolation in the context of glaciation being similar to that of lakes (ibid).

For many species, divergence is clearly related to effects of glaciation, exhibiting multiple lineages derived in as little as 0.1-0.5 million years and exhibiting as little as 0.2-1.5% divergence, upwards to 3.9-7.9% divergence in as little as 0.5-3.5 my (Hewitt 2004; Hershler et al 2005).

The evolutionary history of *M. lustrica* can be inferred using genetic analysis, and this same analysis can be used to infer the relationship between the snail and the landscape. In general, common phylogeographic patterns in glaciated areas include a recolonization pattern whereby population haplotype diversity decreases as the species colonizes previously glaciated territories, and therefore the origin of a particular species can be inferred based on areas of high haplotype diversity (DeChaine and Martin 2004; Rowe et al 2004). More diversity in southern populations, but with more restricted geographies, and more purity in northern populations covering a greater area are expected (Emerson and Hewitt 2005). The population of large regions with multiple unique haplotypes, “by one or two bp” is a clear signal of rapid recent expansion, and can generally be explained by surviving multiple glaciations (ibid). *M. lustrica* likely represents a combination of land based processes of range expansion and contraction, and incorporates some aspects of island processes of isolation.

CHAPTER 3

METHODOLOGY

This project required a number of different approaches, including ecological field surveys, taxonomy and morphology, biogeography, genetic analysis, landscape ecology, and the historical analysis of glaciation.

The collection and habitat assessment included the regional work completed in 1999 (Coote and Roeder 1999) and 2000 (Roeder and Coote 2000), but also included collection of specimens across the range of *M. lustrica* from Massachusetts to Minnesota. Two full-range surveys were completed in 2007 and 2008. The taxonomic and morphological work included extensive specimen evaluation of *M. lustrica* as well as other species, photographic documentation of the various shell forms, analysis of museum lots, and dissection.

The biogeography work included review of historical records of occurrence, investigation of unlikely records outside of the accepted range, and research on the distribution of other organisms during post-glaciation including other invertebrates, fishes, and reptiles. Genetic analysis included the sequencing of two mitochondrial genes for every contemporary population surveyed; analysis of those sequences was used to construct phylogenetic trees and haplotype networks, including sequences of other *Marstonia* spp.

Landscape analysis involves correlating the presence of the various populations of *M. lustrica* with watersheds and glacial patterns across its range at two levels: macro or

the full range, and meso (within) regional watersheds. GIS was used to record field data, develop maps, and for distance measures.

Field Work Protocol

Because of the rarity of this snail, combined with the need for efficient sampling, the survey protocol was designed to cover the greatest range possible while having some certainty of finding populations. The framework used to meet the above requirements was the USGS regional watershed system. Each regional watershed was then broken down by historic population clusters (cells). The use of watersheds is consistent with the potential for gene exchange and isolation, while the use of cells representing historic clusters of populations increased chances for finding the snail. All sites chosen for sampling either were historical records for the snail, or were in close proximity to such sites.

Across the range of *M. lustrica* there are five regional watersheds, each of which is made up of sub-regions, which are further delineated into accounting units, and then finally into catalogue units (USGS 2009). The five regions are: the Upper Mississippi, the Ohio, the Great Lakes, the Mid-Atlantic, and New England. To answer questions at the macro and meso scale, regions were used, as well as a created designation in the form of cells (Fig. 5).

The sampling protocol combined the regional watersheds with nine cells (Fig. 5). Cells were defined as population clusters of *M. lustrica* within a regional watershed, which may cross sub-regional watershed boundaries. The Upper Mississippi was separated into two cells, numbers 1 and 2, the Great Lakes into four cells, numbers 3, 4,

5, and 6, the Ohio into cell number 7, the Mid-Atlantic into cell number 8, and New England into cell number 9. Cell 5 is in Canada and is not part of the USGS system, but is part of the Great Lakes drainage and adjacent to USGS sub-region # 0415. Not

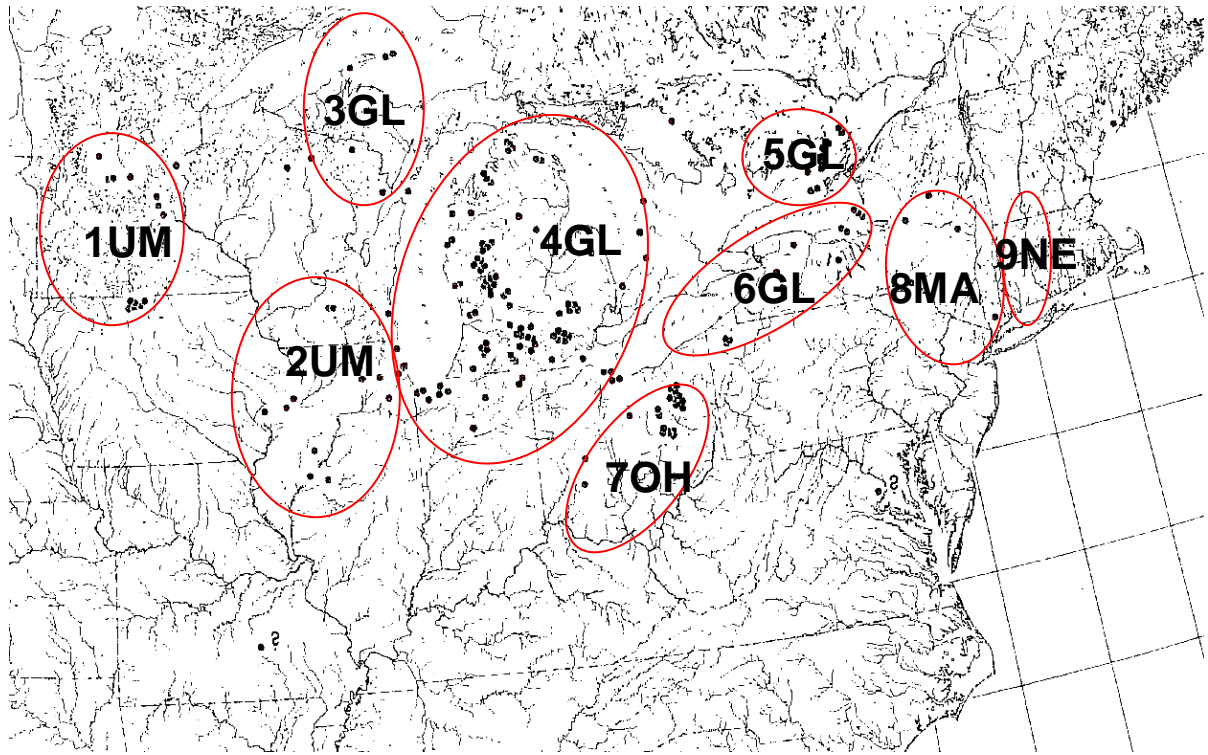


Figure 5 - Map of cells. Distribution of *M. lustrica* showing sampling cells and historical populations of *M. lustrica* (base map Thompson 1977). There are a few populations that do not fall within the designated cells. Watersheds are indicated as follows: Um= Upper Mississippi, GL= Great Lakes, OH= Ohio, MA= Mid-Atlantic, NE= New England.

every historical population of *M. lustrica* was included in the nine cells due to a combination of being exceptionally far from clusters within its own regional watershed, or in close proximity to a cluster in another regional watershed.

I attempted to sample at least three populations within each cell. Sub-regions containing populations of the snail were selected randomly for sampling using the sub-region numbers and a random numbers table. In order to test for genetic variation within

and amongst populations, efforts were made to obtain three or more specimens from each population. In some cases sampling did not result in three specimens.

Snails were collected from each location using a D-net with a 30-second sweep of the vegetation or substrate (Jokinen 1992) in less than 2 m depth, covering approximately 1 m². Specimens for genetic analysis were preserved in the field in 95% ethyl alcohol, while specimens for morphological analysis were anesthetized with menthol before preservation (Jokinen 1992; Hershler pers. comm.). When possible, samples were sorted with a dissection scope in the field. Unfortunately, many samples had substantial mud and debris which made field sorting logistically prohibitive. In those circumstances, whole samples were bottled in water on site to be sorted later. Representative specimens of *M. lustrica* from each site will be archived with Dr. Robert Hershler at the National Museum of Natural History.

Because water chemistry at the watershed level may affect snail distribution (Jokinen 1992), attempts were made to gather that data. Water quality parameters measured included dissolved oxygen, temperature, pH, conductivity, Ca⁺⁺, Mg⁺⁺, Na⁺⁺, and K⁺ (Jokinen 1992). Water quality analyses were done using: a YSI dissolved oxygen meter, an Orion pH meter, a YSI SCT meter, with specific ion tests completed with LaMotte titration kits. Water samples from each location were not always possible (equipment malfunction, etc.).

GIS data were recorded for each location. These data included sample site coordinates, site description, elevation, and any relevant landscape features (e.g., canals). These data was then mapped along with the genetic data, watershed boundaries, and glacial data using ESRI GIS.

In 2000 and 2005, 24 lakes and rivers were sampled in NY, CT, and MA (Appendix A). In 2007 and 2008, an additional 40 lakes and rivers were sampled across the range of the snail (Appendix B), including 18 samples taken along the length of the Hudson River from the Tappan Zee Bridge (above the salt wedge) to Albany, NY. Collectively, 64 water bodies were surveyed for *M. lustrica* from Massachusetts to Minnesota, and from Ohio to Canada.

Materials, lab space, and field equipment required for the study were provided by Bard College at Simon's Rock, with the exception of sequence extraction which was contracted with Dr. Hsiu-Ping Liu (Metropolitan State College of Denver). In addition, sampling of the Hudson River was coordinated through Dr. David Strayer of the Cary Institute. Generally permits are not required to collect invertebrates in the United States, although a permit was acquired in Massachusetts where *M. lustrica* is listed. A permit for collection on Isle Royale, a national park and international bio-reserve, was also obtained, as was one for collecting in Canada.

Taxonomy and Morphology

Despite the difficulties associated with the identification of *M. lustrica*, there are some shell characteristics that I find useful for differentiating *Marstonia* from most other hydrobiids (Table 4). Several researchers have used both male and female reproductive organs for identification of *Marstonia* (Thompson 1977; Hershler and Thompson 1987; Hershler 1994; Hershler and Ponder 1998), but here I rely on qualitative and quantitative shell morphology, mostly ignore soft body parts, and confirm identification using genetic analysis (Wilke et al 2002). In general, the shell of *M. lustrica* is conical with 4-5 whorls,

growing 3.0-6.0 mm long, with colors ranging from translucent beige to brown/black (Berry 1943; Harman and Berg 1971; Jokinen 1992). The whorls are round,

Table 4 – List of morphological character states.

Character	State
Quantitative:	
# of whorls	4-5
Length of shell	3-6 mm
Width of shell	1/3 length
Diameter of aperture	1/4-1/3 length
Qualitative:	
Aperture	Round & detached
Whorls	Moderately shouldered/ rounded
Color	Translucent beige to black

moderately shouldered, with moderately deep sutures, and with the width of the body whorl about 1/3 the length of the shell. The aperture is round, diameter about 1/4 of the shell length, and detached or only slightly touching the body whorl. Species which may be confused with *M. lustrica* using the above character states, with similar ranges, include *M. letsoni*, *F. winkleyi*, *A. limosa*, *Hoyia sheldoni*, *Littoridinops tenuipes* and *Pomatiopsis lapidaria*. However, I believe that with close inspection and side-by-side comparisons, positive identification of these species is possible based on shell morphology alone given enough experience. All snails collected were preserved in full strength, 95% analytic grade ethanol for later identification and possible genetic analysis (Liu pers. comm.). Snails for morphological analysis were anesthetized before fixing using menthol, excluding those specimens for genetic analysis. Snails were identified in the lab to genus or species using standard references (Pilsbry 1890, Harman and Berg 1971; Burch 1980, 1989; Jokinen 1992; Smith 1995).

Phylogenetics

Extraction, PCR, and Sequencing

DNA extraction, PCR, and sequencing were contracted with Dr. Hsiu-Ping Liu. Dr. Liu is a geneticist, having worked extensively with Dr. Hershler and others on the systematics and biogeography of a number of hydrobiidae (Hershler and Liu 2004a & b; Hershler et al 2003 & b, 2007a & b, & 2008; Liu & Hershler 2005 & 2007; Liu et al 2003). Dr. Liu completed the majority of extractions, PCR, and sequencing according to the protocols used in Liu and Hershler (2005). During a visit to her lab in January 2009, she and I extracted DNA and ran PCR on about 20% of the specimens. All sequences reported here were initially edited by Dr. Liu using Sequencer (Gene Codes).

In the case of *Marstonia*, some genetic work was completed previously articulating its basic phylogeny (Hershler et al 2003) and clarifying the anatomical work previously relied upon (Thompson 1977; Hershler and Thompson 1987; Hershler 1994). Here I expand on these previous works, sequencing two mitochondrial DNA markers, cytochrome c oxidase I (COI) and NADH dehydrogenase I (NDI) to infer the historical radiation and phylogeny of *M. lustrica*, and to identify likely isolated populations.

One of the most common genes used for analysis at the species level in gastropods is the mitochondrial gene COI (Wilke et al 2001; Hershler et al 2003; Campbell 2006). This protein coding sequence is viewed as relatively stable and is therefore a common choice for differentiating species. Because each population of *M. lustrica* is essentially confined within a single water body, resulting in relatively little genetic exchange among populations, the use of COI was deemed appropriate for the purpose of identifying trends in genetic structure across the range. Because COI does not

necessarily differentiate closely related populations (e.g., within the same watershed), the more variable mitochondrial marker NDI was used for greater resolution. NDI is a relatively new gene used for mollusks and is considered appropriate for finer scale analysis within regional geographies and may be useful among populations (Liu pers. comm.). Initially, attempts were made to sequence rRNA ITS1 for finer scale analysis, but these efforts were unsuccessful. The employment of COI and NDI takes advantage of the previous work completed on *Marstonia* and related snails (Hershler et al 2007a & b; Hershler et al 2004 a, b; Hershler et al 2003 a, b; Liu et al 2003; Wilke and Davis 2000; Wilke et al 2001), and helps to clarify the phylogeny of *M. lustrica* at the macro and meso levels.

Total genomic DNA was extracted (CTAB protocol) and the respective markers amplified (PCR) by Dr. Liu according to methods articulated in Liu and Hershler (2005). Forward and reverse sequences for COI and NDI were obtained by Dr. Liu using Sequencher 3.1.1 Gene Codes). I completed all secondary edits, alignments, and phylogenetic analysis using MEGA4 (Tamura et al 2008), DNA Alignment 1.3.1.1 (Fluxus 2010), Network 4.5 (Fluxus 2010), and Arlequin 3.1 (Excoffier et al 2005). There were no insertions or deletions for the separate data sets, making for straightforward alignment. Multiple alignment in ClustalW was set to 10.0 for gap opening penalty and 1.0 for gap extension penalty (MEGA4). Sequences were analyzed using pairwise deletion, all three codon positions, and the substitution model was maximum composite likelihood, with 10,000 bootstraps (Hall 2008). Molecular clock was set at 1.7% sequence divergence uncorrected (Wilke et al 2009, Hershler et al 2008). In addition to the independent data sets of COI and NDI, analysis was completed on the two data sets

concatenated. Samples which were only sequenced for one of the two genes were eliminated in the combined data set.

After the construction of trees, the data sets were imported to DNA Alignment 1.3 (Fluxus) and transformed into rdf files for use in Network 4.5. Within DNA Alignment 1.3 sequences were further edited to remove one sample (Long Lake Ohio), which was exceptionally short resulting in an erroneous branch within the resulting COI networks. Sequences were analyzed as multistate files, using the median joining method (maximum parsimony) and the connection cost distance method of Bandelt et al 1999. Sequences of other species reported here are taken from the National Center for Biotechnology Information (NCBI) website.

ClustalW alignment resulted in the inclusion of 55 individuals for COI and 41 for NDI, with 658 and 544 untrimmed positions respectively (MEGA4). There were no gaps in the original data, and so full sequences were used and pairwise deletion utilized. For the concatenated data set (40 individuals), there were 1202 bp, and the sequences translated into protein sequences prior to alignment, using ClustalW with a gap opening penalty of 3 and a gap extension penalty of 1.8 (Hall 2008).

Evolutionary relationships were inferred using the Neighbor Joining method (Jukes-Cantor distance = 0.0154), 10,000 bootstraps, pairwise deletion, all three codon and non-coding positions, and the maximum composite likelihood model (Hall 2008). Changing the above options did not significantly affect the resulting trees. Trees reported here were constructed in MEGA4.

Aligned sequences were loaded into Fluxus Engineering's DNA Alignment program for conversion and use in their Network 4.5 program for the construction of

haplotype networks. Statistical analysis was completed using Arlequin 3.1 (Excoffier et al 2005). Results are presented in three parts: phylogenetic trees, haplotype networks, and statistical analysis.

Quantitative Analysis

Once sequences are obtained, there are two elements to determining and testing genetic patterns across the landscape: phylogenetic and statistical sequence analysis (Manel et al 2003). Phylogenetic analysis, as discussed above, infers patterns in the genetic structure of the sampled populations. The second aspect of analysis is the quantitative or statistical analysis of the sequences. Here I use the analysis of molecular variation (AMOVA) and the Mantel test (Excoffier et al 1992; Manel et al 2003) to examine the genetic structure of the populations across the landscape. Imposed population structures include lake (population), state, USGS regional watershed, and cell. In addition there are three imposed groups (Table 5) that reflect population clusters, regional watersheds, and glacial history.

The Mantel tests for patterns of isolation-by-distance and AMOVAs test for partial sequence variation based on the imposed population structures. Both tests use a modified F statistic for genetic distance (Excoffier et al 1992). The Mantel is run primarily to test the null hypothesis that distance alone will structure populations genetically. A partial Mantel combines the effect of distance with the imposed population structures and genetic distance. The AMOVAs examine the degree to which overall sequence variability is driven by within-population differences or by among-population differences.

Accordingly, Mantel tests were run on Euclidian distance and F_{st} for both COI and NDI, as well as on the population structures that performed the best under the AMOVA analysis. Partial Mantel test were also run on these leading structures using a presence/absence matrix for each population and structure, correlating presence/absence within a structure with F_{st} and Euclidian distance. Because sequence data for NDI are missing from Young Lake NY, the regional groups #1 and #2 share identical structures.

Seven levels of structuring were tested by AMOVA analysis: population (lake), state, USGS regional watershed, cell assignment, and non-watershed regional groups #1, #2, and #3 (Table 5). Non-watershed groupings (Groups #1-3) were developed based on general clustering of populations without regard for watershed boundaries, with the primary differences being variations in the Northeastern and Eastern groupings (Table 5). The variations in the eastern groupings reflect the close proximity of the eastern NY (Hudson River watershed) and western Massachusetts populations (Housatonic River watershed), the clustering of populations in southern Ontario, and the geographically contiguous but isolated population of Young Lake NY which can be placed geographically in either western or eastern NY, or proximal to southern Ontario. The interest in these eastern populations is related to the central question of the origins of the western MA populations.

The results of these tests, in conjunction with the genetic results, are used to illustrate geographically the relationships among populations. The end product is a comprehensive map detailing significant factors in the distribution and genetic makeup of *M. lustrica* across its range, and through time. Using ArcGIS (ESRI 9.3.1), samples were

mapped, along with major water bodies, and dominant haplotypes. Hypothesized glacial advances and associated time periods were also mapped showing the overlay of contemporary populations with those historical features. Phylogenetic analysis was interpreted and mapped out geographically using time frames based on a molecular clock rate of 1.7%/myr.

Table 5 – Regional cluster assignments for AMOVAs.

Group #1	Northwest (MN & Lake Superior), Midwest (MI, OH, & western NY), Northeast (Canada), East (eastern NY & MA)
Group #2	Northwest (MN & Lake Superior), Midwest (MI, OH, & western NY), Northeast (Canada, & Young Lake NY), East (Saw Kill NY & MA)
Group #3	Northwest (MN & Lake Superior), Midwest (MI & OH), Northeast (Canada & western NY), East (Saw Kill NY & MA)

CHAPTER 4

RESULTS

Of the 64 water bodies sampled for the study, a total of 20 populations of *Marstonia* sp. have been confirmed (Appendix E). Ecologically, the lakes sampled represented a wide range of systems, including rivers, streams, large glacial lakes, small glacial potholes, and other natural and man-made systems. The habitats of the particular sample locations varied greatly, from mud to hard rock substrates, from thick submerged vegetation with multiple species including *Vallesineria* sp., *Chara* sp., *Potamogeton* sp. and emergent and floating species to systems with no vegetation. *M. lustrica* was found in all types of habitats, frequently without previously noted presumed obligate species such as *Chara* sp. or *A. limosa*.

Multiple regression analysis using STATISTICA (Statsoft 1999) was completed on the water quality parameters recorded for 27 of the lakes sampled with the number of specimens collected as the dependent variable (Table 6). Results indicated there are no water quality associations between lakes containing *M. lustrica* and those that did not (Adj. $R^2 = 0.012$, $p = 0.44$).

Morphology

As expected, the cryptic nature of the shells and verges was evident within and between populations. However, pictures of shells and verges (male reproductive organ) are included for select populations and discussed to the extent possible, including illustrations of verges from Hershler's work on the genus *Pyrgulopsis* (now *Marstonia*)

(1994) (Figs. 7-11). These pictures demonstrate that there are possibly subtle patterns among *Marstonia* but that overall morphological characters are problematic. It must be

Table 6 - List of lakes and water quality results. Variables are temperature, conductivity, pH, dissolved oxygen, calcium, sodium, magnesium, and potassium, respectively.

location	<i>M. lust</i>	temp	cond	pH	DO	Ca	Na	Mg	K
Catatonk Cr., Tioga, NY	1	22.2	290	8.1	6.1	85	53.3	29	0
Cayuga Lake, Tompkins, NY	0	22.6	390	8.3	6.9	82	8	48	0
Grand Island, Erie, NY	8	26.2	250	8.92	12.2	62	9.2	38	0
Mud Lake, Summit, OH	7	23.4	700	7.72	4.27	118	23	62	0
Long Lake, Summit, OH	55	23.1	790	8.1	6.1	118	23	82	0
Cedar Creek, Champaign, OH	0	19	690	8.0	7.4	260	9.2	90	0
Illinois River, Tazwell, IL	0	27	750	8.2	4.3	90	0	200	35
Pine Creek, Muscatine , IO	0	27	1050	7.9	5.3	44	2.3	226	0
Mississippi River, Rock Island, IL	0	28.5	470	7.7	4.6	38	0	152	0
Eagle Lake, Blue Earth, MN	0	29.5	33	8.7	8.7	60	70	70	0
Harriet Lake, Hennepin, MN	50	29.6	400	8.2	7.8	74	65.3	34	0
St. Croix River, Pine, MN	0	28.3	250	8.1	7.1	56	18.4	34	0
Keeweenaw Bay, Baraga, MI	5	21	110	8.2	7.9	35	0	25	0
Lake Ottawa, Iron River, MI	0	25.7	225	7.9	6.0	64	6.9	51	0
Lake Michigan, Ludington, MI	0	23.7	350	7.6	6.1	120	3.68	42	0
Blue Lake, Muskegon, MI	40	27.2	240	6.8	6.1	60	5.9	52	0
Saint Clair River, Monroe, MI	1	25.2	220	8.0	7.1	72	8.7	44	0
Camden Lake, Moscow, Ontario	8	14.9	300	8.6	3.1	144	18.4	16	10
Lake Ontario, Waupoos, Ontario	6	16.9	250	8.4	6.8	120	24.8	36	0
Moir Lake, Medoc, Ontario	18	17.6	350	8.6	2.8	116	95.7	16	20
Winona Lake, Kosciusko, IN	0	26.2	450	7.45	5.2	120	27.6	40	0
Tippecanoe River, Pulaski, IN	1	22.5	510	8.07	6.4	238	0	32	0
Wampum Lake, Cook, IL	0	28.8	440	8.5	5.1	110	20.7	25	0
Lake Michigan, Manitowoc, WI	0	19.6	365	8.37	10.8	164	24.4	28	0
Greenbay, Brown, WI	0	24.2	410	8.2	4.2	100	9.2	80	0
Limestone Lake, Wright, MN	15	24.5	330	8.25	5.8	80	14.3	84	0
Waverly Lake, Wright, MN	0	26.1	485	7.68	4.07	120	46.9	68	0

noted that specimens dissected in this study are not the same specimens sequenced (due to anesthetization), and since more than one haplotype is present for all of those

populations represented (with the exception of LLMN), correlation between the verges, shell structure, and haplotype is done with caution.

Photographs of verges from this study (courtesy of Dr. Hershler), representing populations identified as distantly related to *M. lustrica* in the phylogenetic analysis, are generally consistent with the genus overall. Interestingly, *M. lustrica* as illustrated in 1994 (Fig. 8-f) does not appear to share any clear resemblance to these specimens (see Fig. 6 for nomenclature). In particular, *M. lustrica* is described as having a “filament short, stubby, without taper; lobe shorter than filament, oblique. Terminal gland elongate, often transverse, borne along distal edge... (Hershler 1994)”. This can be contrasted with *M. letsoni* (Fig. 8-e) which is described as having a “filament medium length, gently tapered; lobe about as long as filament, broad, strongly oblique. Terminal gland small, narrow, transverse, borne along distal edge of lobe (ibid.)”. All of the specimens presented from this study appear to resemble *Marstonia* spp. other than *M. lustrica*.

The greatest similarity between the illustrations include *M. arga* and LLMN & HLMN, *M. castor*, *M. halcyon*, and *M. letsoni* with IRMI, KBMI, and LOCA, and possibly *M. pachyta* with MLCA. While *M. scaliformis* is included as part of the genus, it is ignored here as that species has a highly distinctive shell.

The comparison of these illustrations does not contribute much to the resolution of the phylogeny. For example, the similarity of verge form between *M. castor* and *M. halcyon* with IRMI, KBMI, and LOCA is incongruent with the inferred phylogeny. Most interesting among these illustrations is the verge of *M. arga* (Fig 8-b) and its similarity to LLMN and HLMN. These forms do appear to be relatively distinct and the two populations are both relatively isolated and identified as having distinct haplotypes in the

phylogenetic analysis. Unfortunately there are no published sequences for *M. arga*. In short, the use of verge structure, and considering in part the depauperate data set presented here, does not add to this phylogenetic analysis.

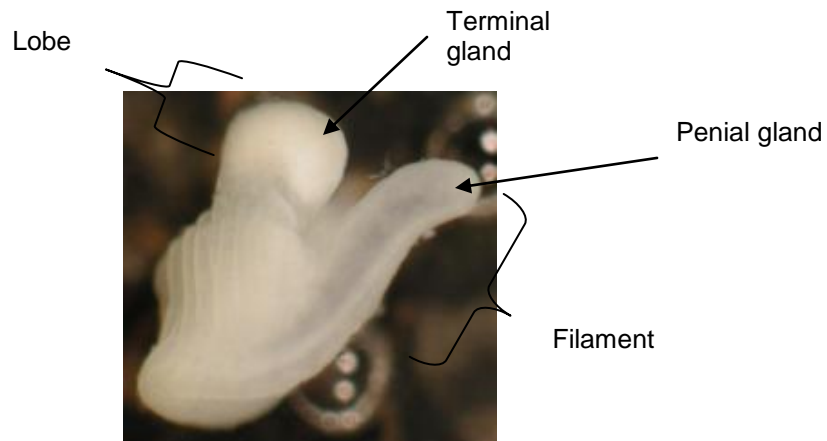


Figure 6 – Nomenclature for verge illustrations.

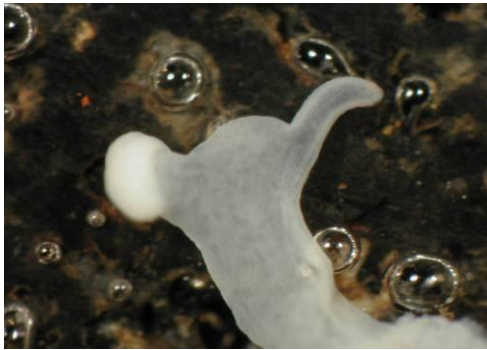
Similar to the illustration of the verges, shell morphology is cryptic as well. Figures 10 and 11 show shells of museum lots for select species and select shells from this study, illustrating the overall similarity within the *Marstonia* genus, as well as within lot variability. In every population sampled where more than one specimen was found, shells varied within the parameters of the character states articulated earlier (Table 4). The shells shown from this study (Fig. 11) represent the range of forms observed. It should be noted that of the museum lots, there are no published sequences for *M. letsoni* or *M. arga*.



Isle Royale, MI (IRMI)



Harriet Lake, MN (HLMN)



Limestone Lake, MN (LLMN)



Keweenaw Bay, MI (KBMI)



Moira Lake, CN (MLCA)



Lake Ontario, CN (LOCA)

Figure 7 – Verges from select populations. Photographs of verges from select populations identified as containing haplotypes distantly related to *M. lustrica*.

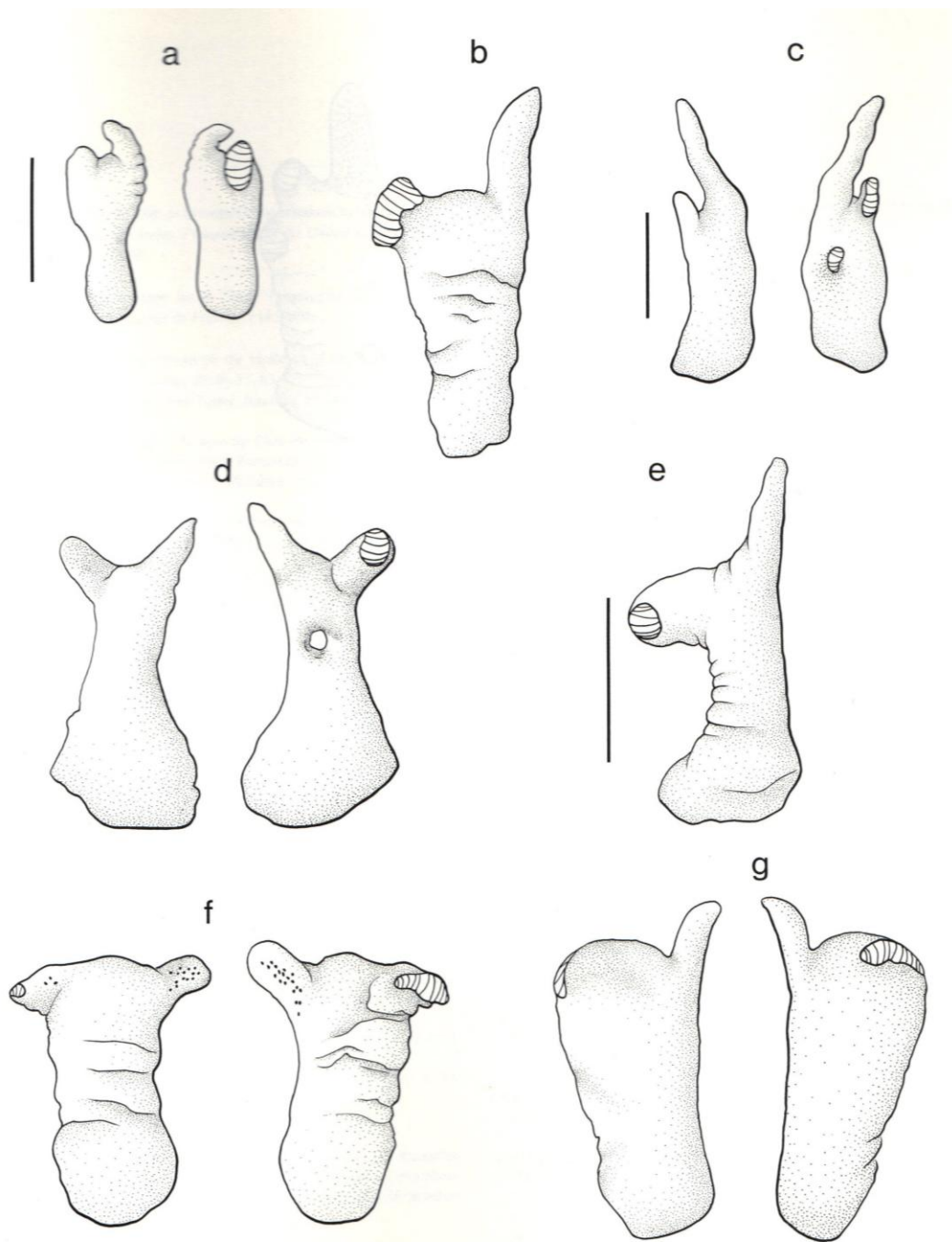


FIGURE 52.—Penes of eastern *Pyrgulopsis*: a, *P. agarhecta*, UF 189468 (bar = 0.5 mm); b, *P. arga*, UF uncat.; c, *P. castor*, UF 189470 (bar = 0.25 mm); d, *P. halcyon*, UF 146698; e, *P. letsoni*, UF 91727 (bar = 0.5 mm); f, *P. lustrica*, UF 22227; g, *P. ogmophaphe*, UF 146697. (Scale for b,d,f,g as for a.)

Figure 8 - Verges from select *Marstonia* spp. A., *M. agarhecta*; b, *M. arga*; c, *M. castor* (bar=0.25 mm); d, *M. halcyon*; e, *M. letsoni* (bar=0.5 mm); f, *M. lustrica* (Clark Co., OH); g, *M. ogmophaphe*. (Scale bar 0.5 mm, b,d,f,g, as for a.) (Hershler 1994).

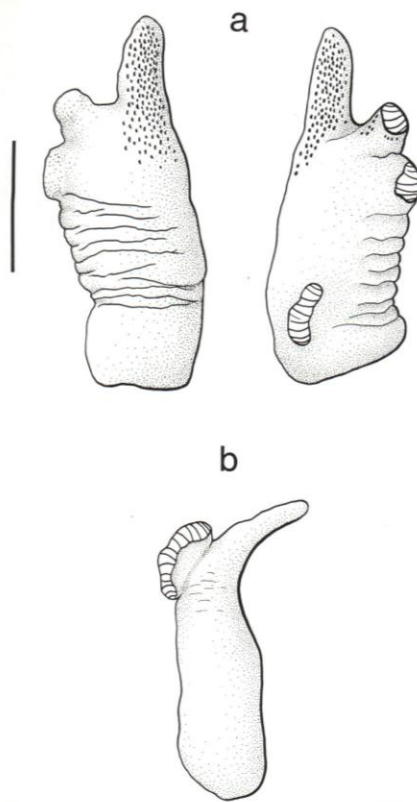
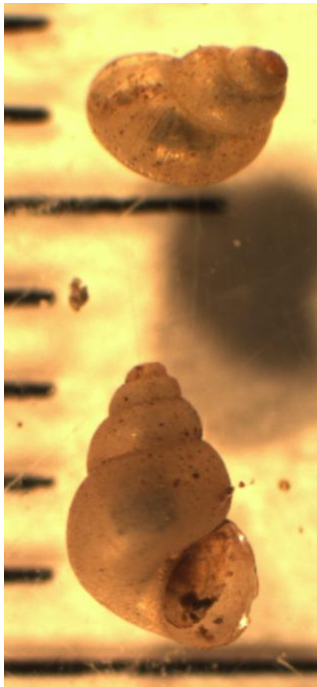
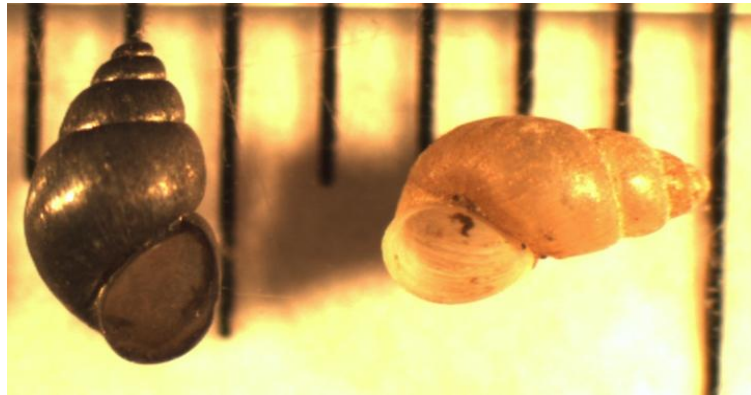


FIGURE 53.—Penes of eastern *Pyrgulopsis*: a, *P. pachyta*, UF 189462 (bar = 0.5 mm); b, *P. scalariformis*, UF 91726 (scale as for a).

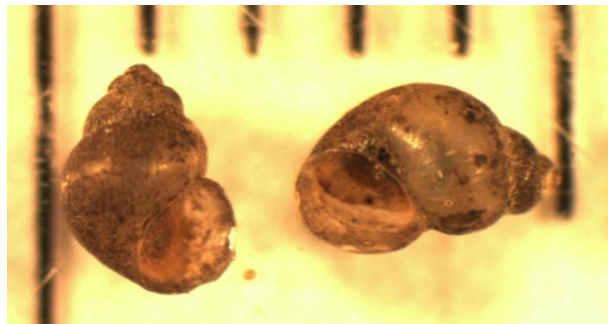
Figure 9 - Verges from select *Marstonia* spp. Cont.: a, *M. pachyta*; b, *M. scalariformis* (bar=0.5 mm) (Hershler 1994).



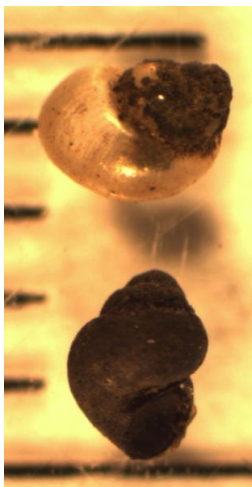
M. arga - Guntersville , AL



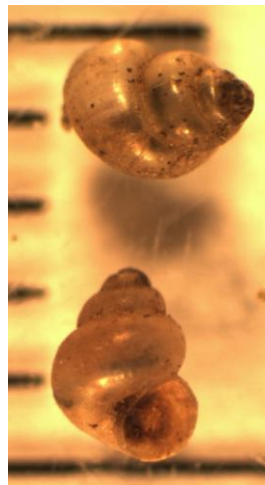
M. letsoni - Crescent Lake, MI



M. pachyta - Limestone Creek, AL

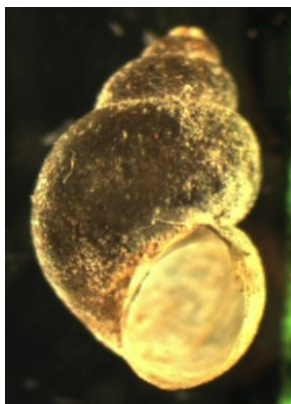


M. halcyon - Ogeechee River, GA



M. agarhecta - Bluff Creek GA

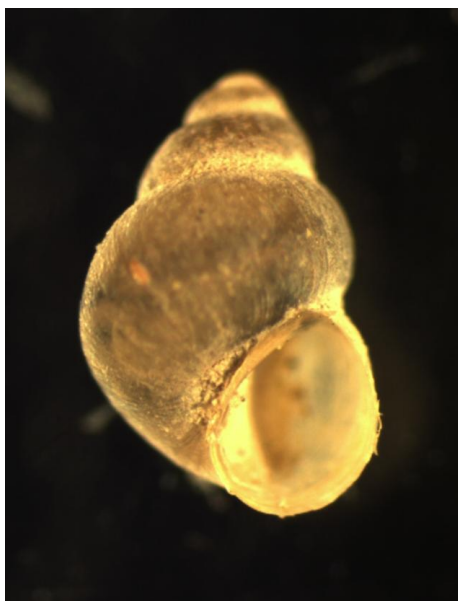
Figure 10 – Shell from select *Marstonia* spp. from the Florida Museum of Natural History and the National Museum of Natural History. Scale is in mm.



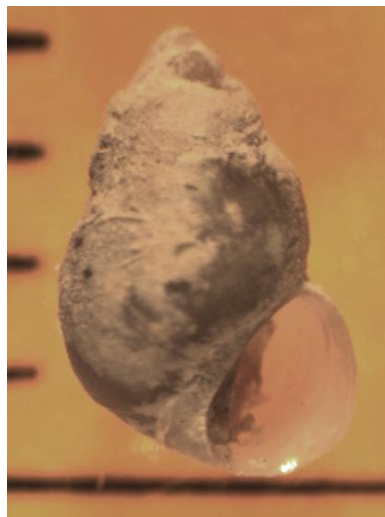
Keweenaw Bay, MI (KBMI)



Limestone Lake, MN (LLMN)



Sinclair River, MI (SRMI)



Lake Huron, MI (LHMI)

Figure 11 – Shells from select populations. Three populations inferred to be distantly related to *M. lustrica* (KBMI, LLMN and LHMI), and one population interpreted as *M. lustrica* (SRMI) based on genetic analysis. All shells are 3-4 mm in length.

Phylogenetics

Specimens of *M. lustrica* from Stockbridge Bowl, MA were sent to Dr. Robert Hershler at the Smithsonian Institute in 2000 for genetic analysis using the mtDNA marker COI and for morphological analysis as part of his work (Hershler et al 2003). The genetic analysis of the Stockbridge Bowl population established the COI haplotype sequence for *M. lustrica* (NCBI # AF520945), and confirmed its placement as sister to *M. pachyta* and *M. comalensis* (ibid.). Additional genetic analysis for the pilot study in 2005 included four populations: Young Lake, NY (type locality), the Saw Kill, NY, Stockbridge Bowl, MA and Laurel Lake, MA. This work was contracted with Dr. Liu, and was consistent with the earlier analysis of Stockbridge Bowl, indicating that the populations in Laurel Lake and Stockbridge Bowl shared identical COI sequences, and that there was a single nucleotide polymorphism (SNP) between the two populations in NY as well as between MA and the New York populations (Table 7). This pilot study suggested that within population variability for COI was minimal but that among population variability was informative, specifically that there was isolation occurring between these four eastern populations.

Based on the pilot study, samples from the 2007 and 2008 field surveys were sent to Dr. Liu for sequencing of COI and NDI. Of the approximately 70 lakes and rivers surveyed from western Massachusetts to Minnesota, 30 sites yielded hydrobiidae snails that were likely of the *Marstonia* genus. The number of specimens from each site submitted for DNA extraction and sequencing ranged from one to five with typically two-three specimens per site. Genetic analysis was completed for COI and NDI. Of the initial specimens submitted for extraction a total of 55 were successfully sequenced for

COI, and 41 were successfully sequenced for NDI. The total number of water bodies represented by COI and NDI is 20 and 17 respectively, and 17 for both.

Table 7 – Haplotype list from pilot study. Single nucleotide polymorphisms (SNP) for COI between four eastern populations of *M. lustrica* from the 2005 study. The within population variation is zero and Stockbridge Bowl and Laurel Lake are identical.

Location	Specimen	SNP:
SBMA	A	GG
SBMA	B	GG
SBMA	C	GG
LLMA	A	GG
LLMA	B	GG
SRNY	A	AA
SRNY	B	AA
YLNy	A	GA
YLNy	B	GA
YLNy	C	GA
YLNy	D	GA

SBMA:Stockbridge Bowl, MA; LLMA:Laurel Lake, MA; SRNY:Saw Kill, NY;
YLNy:Young Lake, NY

Two populations submitted were questionable and suspected to be other species. One was confirmed as a juvenile *A. limosa*, the second from Winona Lake, IN, resulted in an unknown sequence, with the two closest sequences on NCBI being *Cochliopa texana* and *Pyrgulopsis notidicola*. A similar specimen was submitted to Dr. Hershler for morphological analysis and was identified as *Lyogyrus granum* (for which there are no published sequences). The four sub-populations collected from the Hudson River, which were initially identified as *M. lustrica*, were in fact *F. winkleyi*, a species not previously collected in fully freshwater systems and morphologically virtually identical to *M. lustrica*. These specimens were not included in analysis.

Sequence Analysis

The unrooted trees constructed for COI and NDI clearly show similar structures (Figs. 12-14). The primary difference between the two trees is that NDI shows a deeper divergence, in particular, taking MLOHd and BLMlb out of the unresolved clade at the top of the COI tree, and placing them among the more basal clade of HLMNa & b. This particular difference is affirmed in the concatenated tree (Fig. 14). The concatenated tree includes three other species of *Marstonia* for which both COI and NDI sequences are available (NCBI).

By inserting sequences from other species of *Marstonia* into the concatenated dataset, the overall structure of the sampled populations gets clearer and further demonstrates the congruence of the two independent datasets (Fig. 14). Returning to COI only, it is possible to include all of the currently published COI sequences for *Marstonia* spp. (NCBI) and clearly show their relationship to the sampled populations (the unresolved clade at the top of the tree is collapsed for illustration purposes) (Fig. 15). What is clear from this tree is that the clustering of sequences outside of the main *M. lustrica* clade are associated with previously recorded sequences for six other identified *Marstonia* spp. In particular note that the cluster of northwest sequences associated with *M. pachyta* which may or may not be that species but which clearly fall outside of the *M. lustrica* clade, raising the possibility they may be a different species altogether. Also note that the three sequences from GINY (Niagara River, Grand Isle, NY) and TRIN (Tippicanoe River, IN) are deeply divergent from any other the species represented, are most closely related to the southeastern *Marstonia* spp. and most likely are a new species of *Marstonia*.

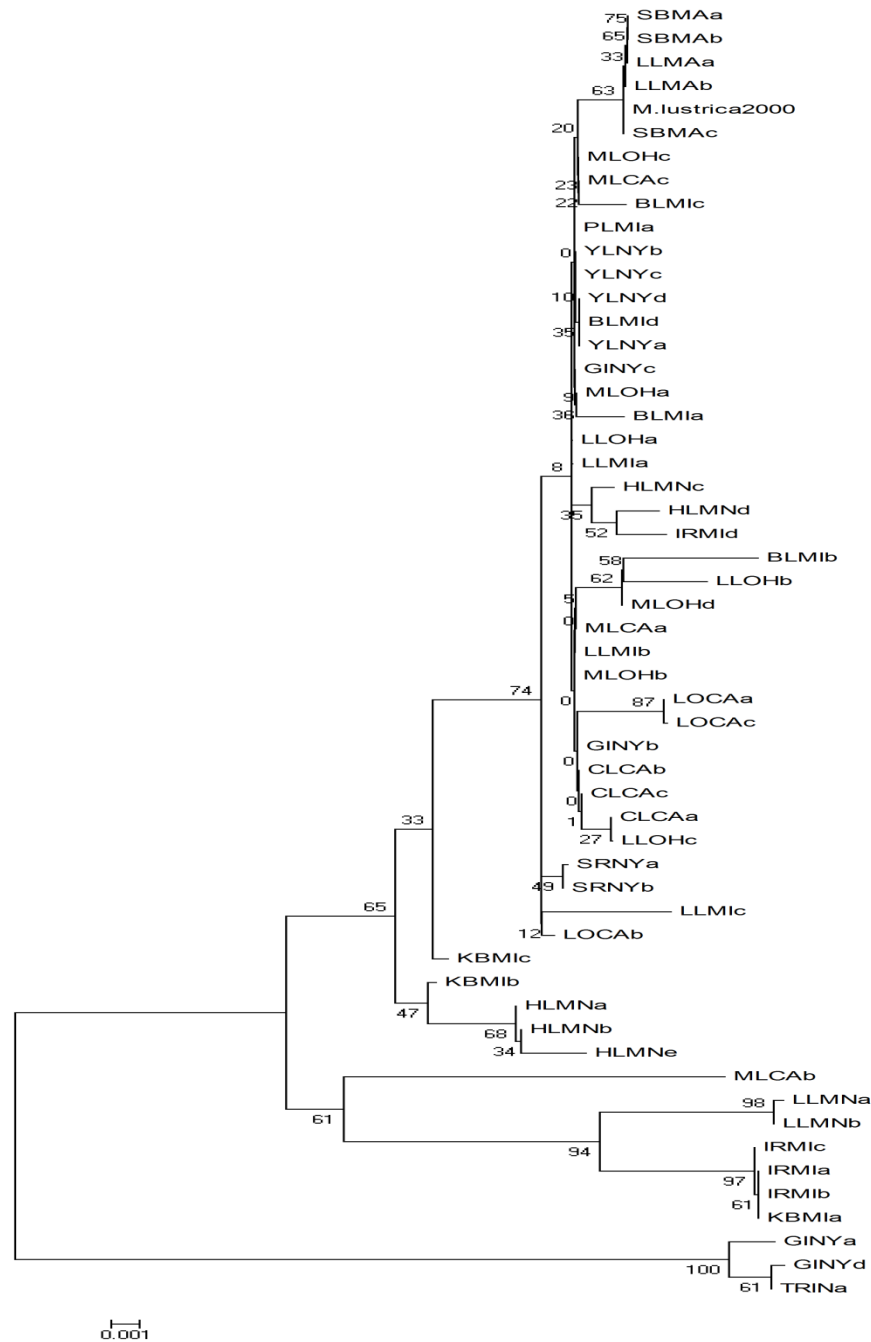


Figure 12 – Neighbor-joining tree for COI. Evolutionary relationships of 55 COI taxa: The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10,000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. There were a total of 658 positions in the final dataset.

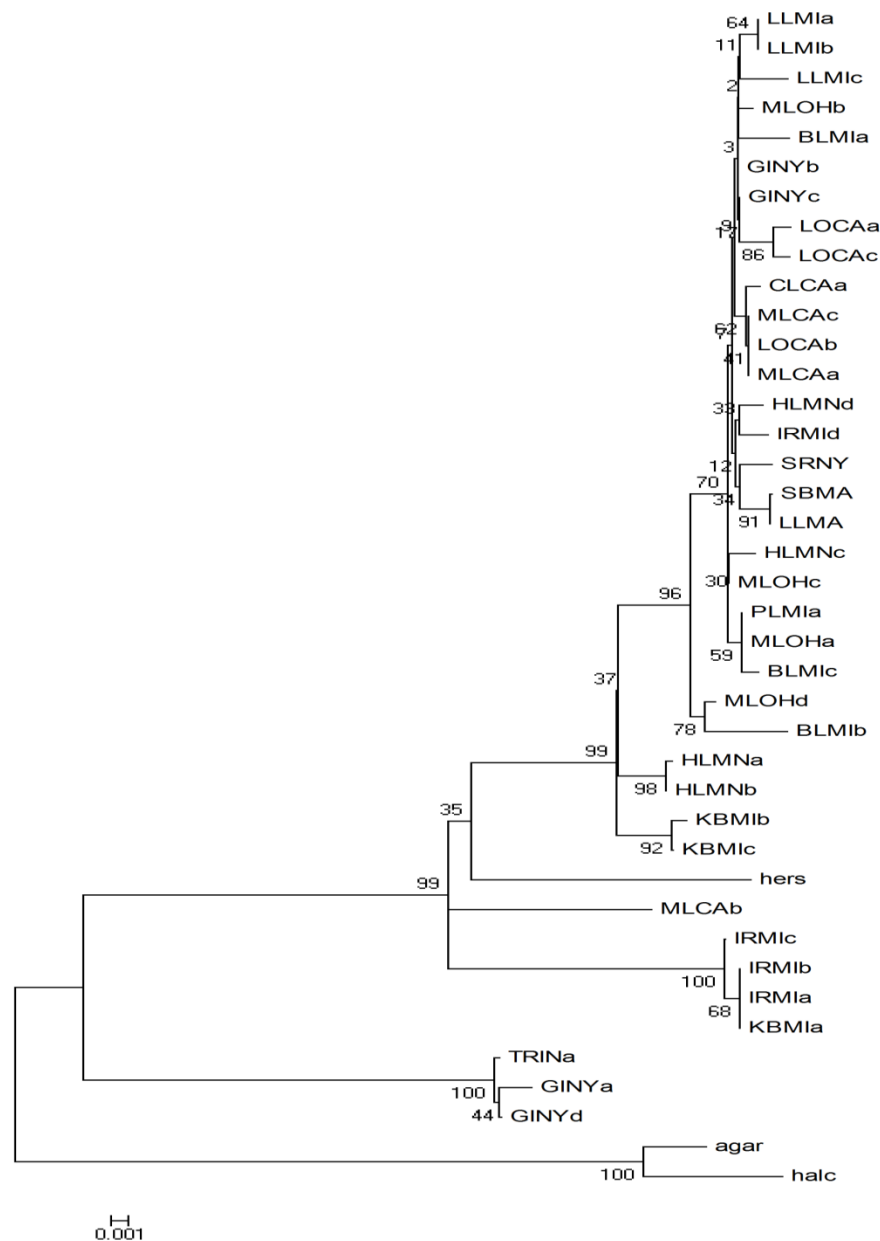


Figure 14 – Neighbor-joining tree for concatenated data. Evolutionary relationships of 40 concatenated taxa, with three outgroups (hersh= *M. hershleri*, agar= *M. agarhecta*, halc= *M. halcyon*): The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10,000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. There were a total of 1202 positions in the final dataset.

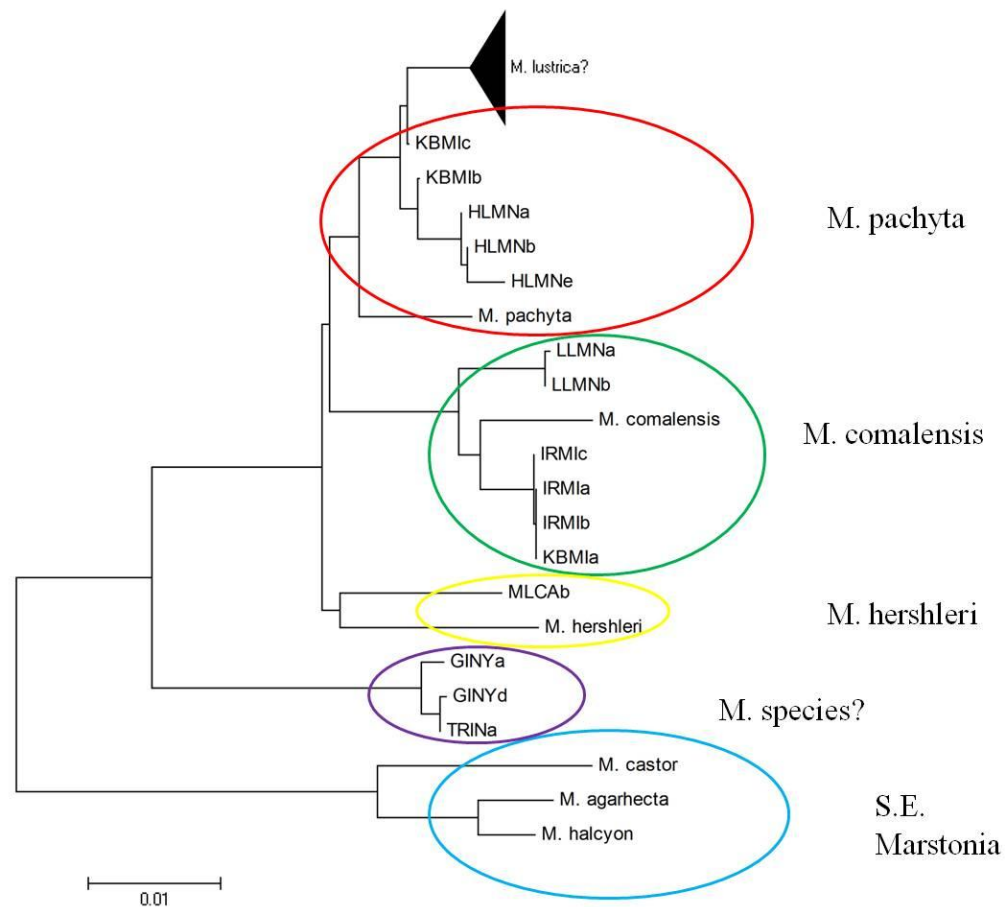


Figure 15 – Collapsed tree for COI with outspecies. Evolutionary relationships of 61 COI taxa including outspecies (no bootstrap) and hypothesized clades (*M. lustrica?* = unresolved taxa collapsed), hypothesized groupings of species listed to the right of the tree. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.22256913 is shown. There were a total of 658 positions in the final dataset.

Haplotype networks represent the same sequence data in a different format.

Network 4.5 only analyses parsimony informative sequence differences, without the option for pairwise deletion, using maximum parsimony to build the shortest trees. Here I have employed the median-joining algorithm in Network 4.5 to develop the shortest

networks. As a result of using only the parsimony informative sites, some of the pairwise variability evident in the phylogenetic trees is lost.

The haplotype networks for COI and NDI (Figs. 16-18) show similar structures, and generally support their respective trees illustrated above. However, there are some significant differences, particularly for SRNY populations. In the COI network it appears that SRNY is an intermediary between the main haplotype(s) and the other species. This is due to the shortening of sequences in Network 4.5 and is not supported in either the NDI or concatenated networks.

Haplotype network results for the concatenated data (Fig. 18) shows a primary haplotype represented by GINYb, present in three midwest populations GINY, MLOH, and PLMI. The next largest grouping LOCAb, represents two populations from Canada, including MLCAa & c. SBMAa represents the two Massachusetts populations (SBMA & LLMA), and LLMIa represents two samples from Long Lake MI. The remaining nodes are single sequences with unique haplotypes (singletons).

The concatenated tree with all published species included (Fig. 19) clearly demonstrates the clustering around outspecies as well as the tight clustering of the unresolved *M. lustrica* clade. It highlights distinct divergence for MLOHb and BLMib within the *M. lustrica* clade, and shows the association of *M. pachyta* with the populations from HLMN and KBMI. It also includes a molecular clock set to 1.7%/myr (Wilke et al 2009).

To further examine the *M. lustrica* clade including BLMib and MLOHd, mismatch analysis in Network and Arlequin was completed. When sequences MLOHb and BLMib are included in the analysis there is a unimodal distribution which is typical

for a recent expansion, and there was no significant difference from the demographic expansion model [$p(ssd) = 0.70$]. However, the F_u 's test was insignificant ($F_s = -0.036$, $p = 0.54$) and does not support a recent expansion. When BLMib and MLOHd are excluded, the mismatch distribution remains unimodal, but with a significant F_u ($F_s = -8.30$, $p = .000$) supporting recent demographic expansion, with no significant difference from the demographic expansion model [$p(ssd) = 0.68$]. These results support a common and recent expansion for all of the clade, with BLMib and MLOHd simply being representative of slightly older populations having emerged during one of the many interglacials. This interpretation is consistent with the view that the lower portion of Michigan and Ohio currently contain populations in closest proximity to the glacial front and that as such these populations likely are derived from an older, more stable source population(s).

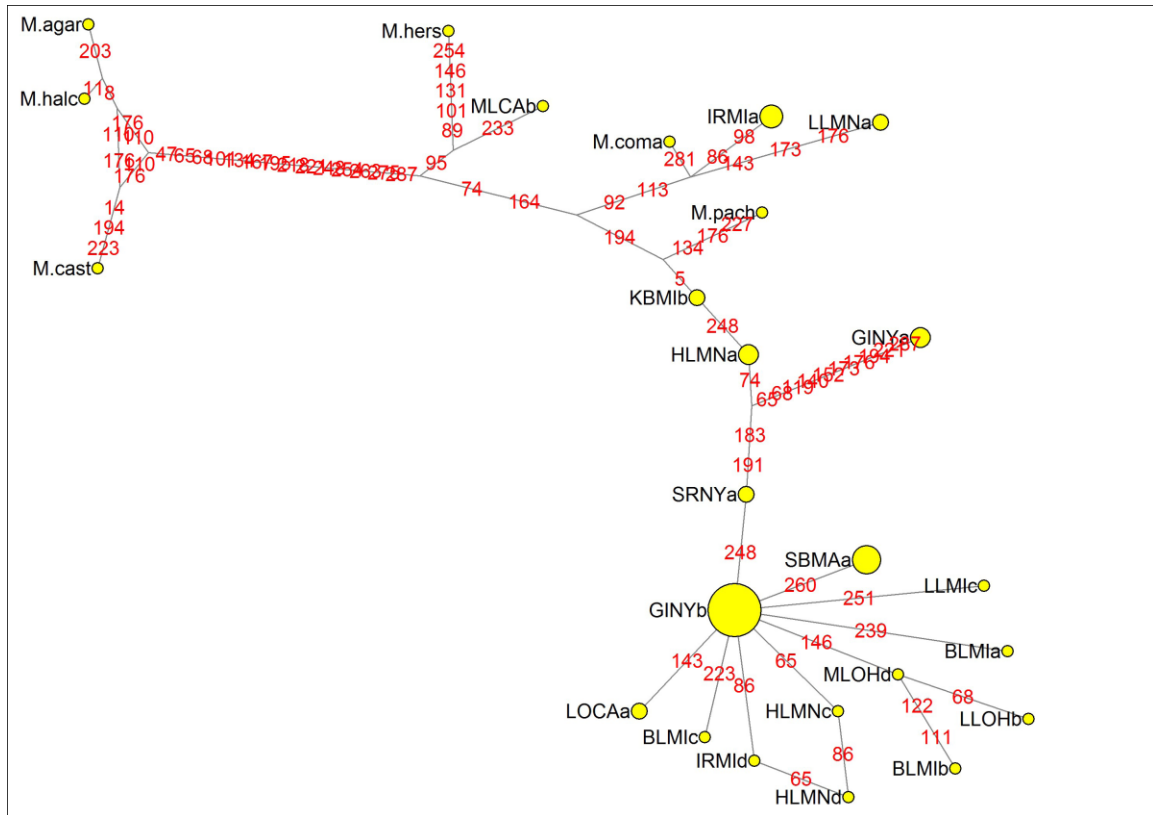


Figure 16 - Haplotype network for COI. The dominant haplotype is GINYb, which includes 21 taxa and represents the Midwest and Canadian populations, but does not include any eastern or western populations. There are 12 other haplotypes surrounding GINYb, differing by 1-3 bp. SBMAa represents all of the Massachusetts taxa. SRNYa appears to be an intermediary for all other samples and species, but this is not supported in the COI tree, the NDI tree or network, or the concatenated tree or network, and reflects a loss of variability in the COI dataset due to editing.

In summary, the unresolved clade of *M. lustrica* is well differentiated from the other samples (Fig. 19), placing several of the specimens collected with other species of *Marstonia*. For COI there is only one derived and wide-spread haplotype. Excluding other species, Blue Lake in central Michigan contains the highest number of COI haplotypes (4), and the Great Lakes drainage contains the highest number of unique concatenated haplotypes with a total of 11 out of 15. Of the four remaining haplotypes,

two are from the upper Mississippi (MN), one is from the Mid-Atlantic drainage (Hudson River), and one is from the New England drainage (MA).

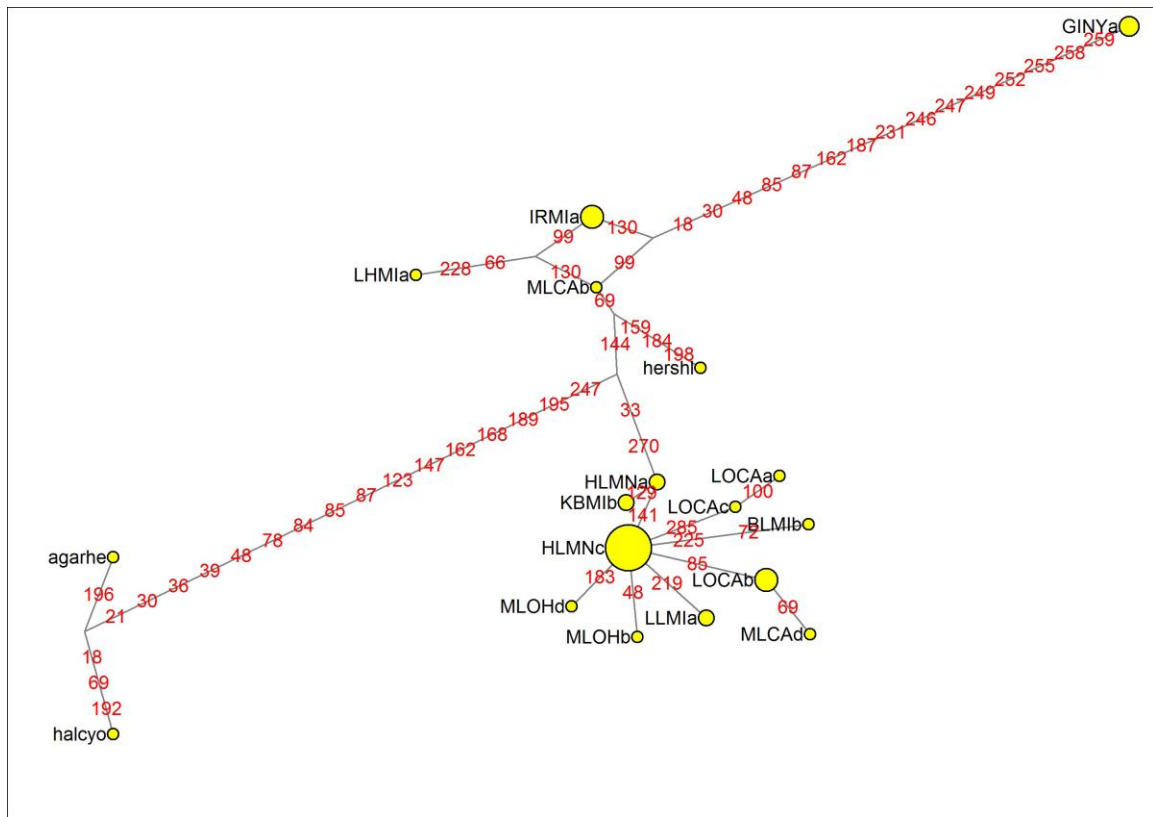


Figure 17- Haplotype network for NDI. The dominant haplotype is HLMNc, which consists of 16 taxa from across the entire range of the study (including SRNY), with the exception of Canada. There are ten additional haplotypes surrounding HLMNc, differing by 1-2 bp.

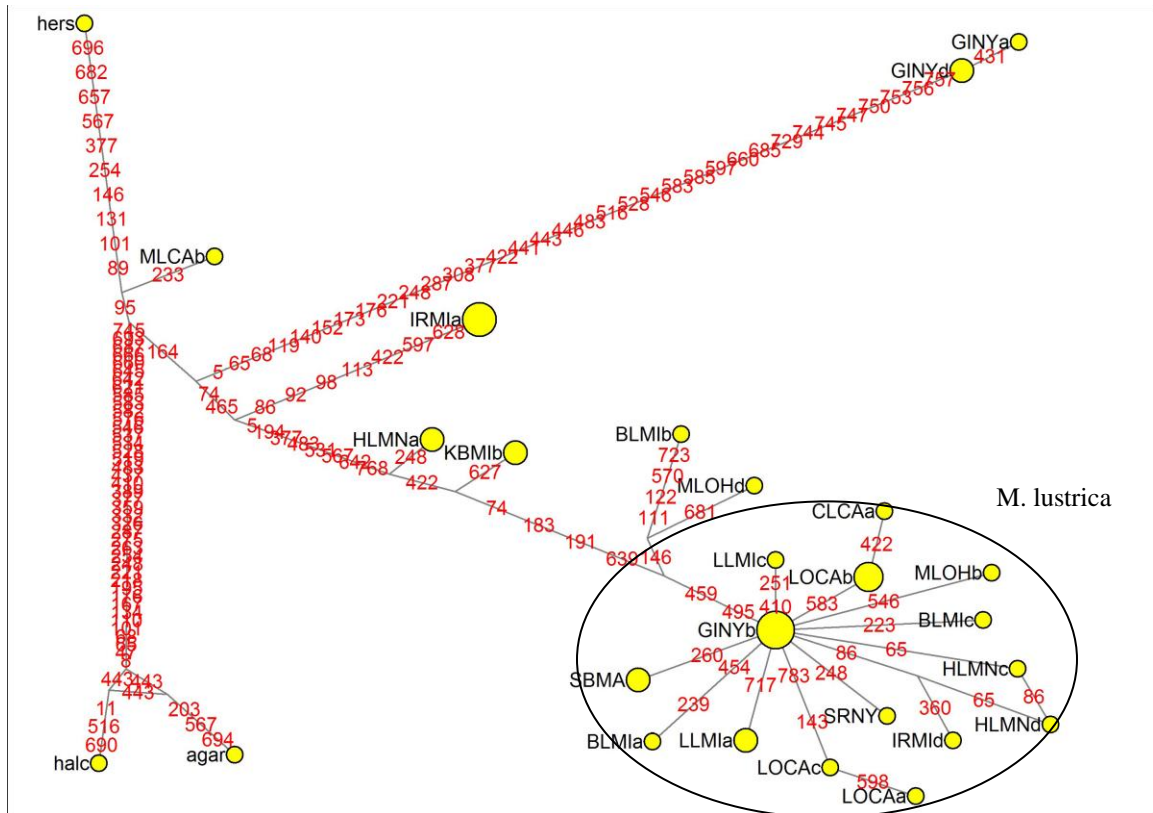


Figure 18 - Haplotype network for concatenated data. Network shown is the original, of the three shortest networks constructed. GINYb represents a total of 5 taxa from the Midwest populations only. LOCAb represents two Canadian populations, SBMA represents the two Massachusetts populations, IRMIa represents the two populations in Lake Superior, and GINYd represents the Niagara River population and the Tippicanoe IN population. All other nodes represent single populations.

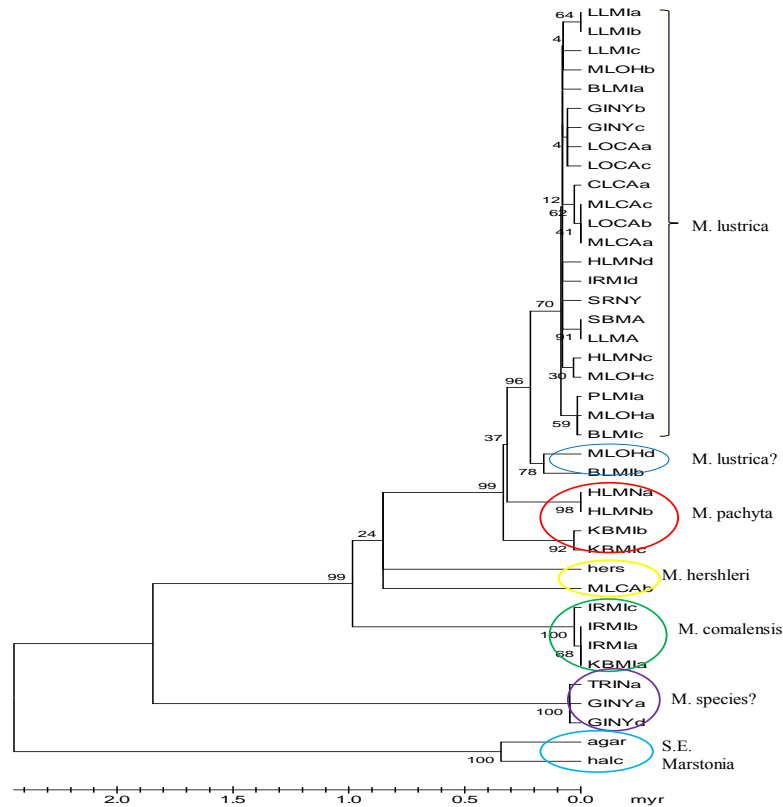


Figure 19 – Phylogenetic tree of concatenated data with hypothesized species associations. Evolutionary relationships of 40 concatenated taxa (linearized) with hypothesized species associations highlighted and labeled, and a molecular clock calibrated to 1.7% (Wilke et al 2009). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10,000 replicates) are shown next to the branches. The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages. The clock calibration to convert distance to time was 58.82 (time/node height). There were a total of 1202 positions in the final dataset.

The dominant COI haplotype is represented across the eastern two thirds of the range, showing up in ten sites: Young Lake NY, Niagara River NY, Lake Ontario Canada, Moira Lake Canada, Camden Lake Canada, Long Lake Michigan, Blue Lake

Michigan, Portage Lake Michigan, Mud Lake Ohio, and Long Lake Ohio. The dominant NDI haplotype is represented across the range, with the exception of Canada.

Within the concatenated *M. lustrica* clade, there are 15 novel haplotypes, with only one of those (GINYb) represented broadly, including midwest and northeastern populations only, suggesting barriers to gene flow to the western and eastern populations.

With regards to the samples not included in the main *M. lustrica* clade, there are some clear patterns in both the COI and NDI trees and networks:

- A. It is clear that the specimen collected at Tippicanoe, IN and the two specimens from the Niagara River (Grand Island), NY are a distinct species, possibly *M. letsoni* which has yet to be sequenced but has been found in the area by other researchers.
- B. The second clear clustering of specimens are from Isle Royale and Keweenaw Bay, MI, Limestone Lake MN, and Lake Huron MI. These specimens are closely related to *M. comalensis*.
- C. There is a third possible species from Moira Lake, Canada which is closely related to *M. hershleri*, but which is also closely related to the specimen from Limestone Lake, MN mentioned above (B).
- D. A fourth cluster occurs with several specimens from Harriet Lake, MN and Keweenaw Bay (Lake Superior), MI. It seems likely that this grouping represents either a new species of *Marstonia*, or is *M. pachyta*.
- E. There is a fifth grouping of haplotypes including specimens from Mud Lake and Long Lake, Ohio and Blue Lake, MI, but it is only differentiated from the

main *M. lustrica* clade by four to six bp. Mud Lake and Blue Lake have stronger support under NDI, and there are no NDI data for Long Lake OH.

Results for independent and concatenated data support the presence of 3-5 species other than *M. lustrica* across the range of this study.

Quantitative Analysis

Overall the AMOVAs demonstrate regional structuring of the populations and the Mantel test for Euclidian distance indicates a weak effect of distance. The AMOVAs illustrate a range of regional effects while the Mantel tests on the same structures were run on the best performing AMOVAs for comparison.

Because it is clear from the phylogenetic analysis that species other than *M. lustrica* were collected, a judgment had to be made on which sequences represented *M. lustrica* and which sequences did not. Accordingly, those sequences located within the main *M. lustrica* clade, and the two sequences located immediately adjacent to the main clade (MLOHd and BLMib) (Fig. 19) are included in this analysis while the more distant sequences are excluded, unless otherwise noted.

The inclusion of those taxa outside of the main clade drives up among-population variability relative to within-population variability, inaccurately quantifying variation at the species level as demonstrated by analysis of COI (Table 8). For reference purposes, all sequences are included in the Mantel test for isolation-by-distance (Tables 9 & 10).

The Mantel test on Euclidian distance (kilometers) and Fst (p-dist) for both COI and NDI are reported in Tables 9 and 10. The Mantel for COI result in a modest but significant $r=0.2770$ ($r^2=0.0767$, $p=0.01$) for all sequences, and a modest but significant $r=0.2337$

Table 8 - AMOVA results for COI sequences. The *M. lustrica* clade only versus an AMOVA on all samples (including likely other species) at the population level. Inclusion of likely outspecies drives up among population variability. Ss=sum of squares, Vc=variance components, % Var=percent variance.

COI		M. lustrica only			All samples		
source of variation		Ss	Vc	% Var	Ss	Vc	% Var
among populations		16.25	0.27	38.03	110.06	1.59	48.61
within populations		10.92	0.44	61.96	61.38	1.68	51.38
Fst		0.38			0.49		
p-value		0.00			0.00		

($r^2=0.0546$, $p=0.05$) for the *M. lustrica* clade only. The Mantel for NDI, which exhibits greater within-population variability, has a weaker relationship, with an $r=0.0428$ ($r^2=0.0018$, $p=0.35$) for all sequences, and an $r=0.1872$ ($r^2=0.0350$, $p=0.70$) for *M. lustrica* only. A Mantel for COI was also run for the eastern populations only (including Canada) resulting in a better distance relationship ($r=0.39$, $r^2=0.15$, $p=.04$) than for the *M. lustrica* clade as a whole (Table 9). In all cases, the r^2 values indicate a weak distance signal.

Results of the AMOVAs (Table 13) show within-population variation for COI (57.28-61.96%) with significant p-values for every structure. For NDI the within-population variation is higher (77.56-89.41%), which is consistent with greater AMOVA results for among-groups for COI ranged from 17.21-38.03%, with significant p-values. NDI's among-groups results range from 10.58-35.41%, with significant p-values for all groupings except at the population level (10.58%, $p=0.16$).

The highest levels of sequence variability among groups for COI are Population (38.03%), Cell (31.28%), USGS regional watershed (27.73%) and Structure 1 (27.06%).

Table 9 – Mantel results for COI and Euclidian distance.

COI	All samples	M. lustrica only	Eastern populations only
R	0.2770	0.2337	0.3886
r ²	0.0767	0.0546	0.1510
P	0.0050	0.0490	0.036

Table 10 – Mantel results for NDI and Euclidian distance.

NDI	All samples	M. lustrica only
R	0.0428	0.1872
r ²	0.0018	0.0350
P	0.353	0.066

The three highest for NDI are USGS regional watershed (35.41%), Cell (31.28%), and State (23.00%). Population level aside, taking COI and NDI together, the highest among-groups variations are at the watershed and cell levels, while Structure 2 and 3 did relatively poorly. Importantly, the variation among populations within groups for all structures is low for both COI and NDI (less than 18%).

Using a presence/absence (1/0) matrix for populations in each given structure, Mantel tests on the two leading structures for COI as indicated by AMOVA do not show any significant relationships between Fst estimates and population structures (Table 11). In other words, the genetic distance as measured by Fst does not breakdown along the lines of either watershed or cell, indicating that the extent of genetic distance within each

structure is consistent among groups. A partial Mantel on these two leading structures for COI, distance, and Fst show no significant partial correlations (Table 12). This would be expected given that neither the Mantel on distance nor the Mantel on presence/absence for the two leading structures are significant.

Table 11 – Mantel results for watershed and cell.

Structure	COI ($r/r^2/p$)	NDI ($r/r^2/p$)
cell	-0.21/0.04/0.99	-0.17/0.03/0.95
watershed	-0.20/0.04/0.90	-0.17/0.03/0.88

Table 12 – Partial Mantel results for watershed and cell.

Structure	r (structure, dist.)	r (dist., structure)
cell	-0.1176 (p=0.91)	0.1615 (p=0.12)
watershed	-0.1493 (p=0.83)	0.1890 (p=0.07)

Table 13 - AMOVA for COI and NDI. Ss=sum of squares, Vc=variance components, % Var=percent variance, Fct, Fsc, and Fst= the proportion of sequence diversity of sub-populations relative to the total. P-value is the probability of obtaining a greater variance component and F statistic than observed: Fct=among groups, Fsc=among populations within groups, Fst=within populations. Structures #1 and #2 for NDI are identical.

Structure	COI				NDI			
Population (Lake)								
source of variation	Ss	Vc	% Var		Ss	Vc	% Var	
among populations	16.25	0.27	38.03		24.14	0.18	10.58	
within populations	10.92	0.44	61.96		16.92	1.53	89.41	
Fst	0.38				0.11			
p-value	0.00				0.16			
Watershed								
source of variation	Ss	Vc	% Var	Fct/Fsc/Fst (p)	Ss	Vc	% Var	Fct/Fsc/Fst (p)
among groups	9.06	0.21	27.73	0.28 (0.01)	13.24	0.70	35.41	0.35 (0.00)
among pop within groups	7.19	0.11	14.51	0.20 (0.10)	10.90	-0.26	-13.3	-0.21 (0.45)
within populations	10.92	0.44	57.75	0.42 (0.00)	16.92	1.53	77.89	0.22 (0.16)
Cell								
source of variation	Ss	Vc	% Var	Fct/Fsc/Fst (p)	Ss	Vc	% Var	Fct/Fsc/Fst (p)
among groups	11.98	0.23	31.28	0.32 (0.00)	17.67	0.56	31.62	0.32 (0.00)
among pop within groups	4.28	0.06	8.54	0.12 (0.27)	6.47	-0.32	-17.99	-0.26 (0.69)
within populations	10.92	0.44	60.18	0.40 (0.00)	16.92	1.53	86.38	0.14 (0.14)
Structure 1								
source of variation	Ss	Vc	% Var	Fct/Fsc/Fst (p)	Ss	Vc	% Var	Fct/Fsc/Fst (p)
among groups	7.71	0.21	27.06	0.27 (0.00)	10.66	0.37	20.48	0.20 (0.01)
among pop within groups	8.55	0.13	16.15	0.22 (0.01)	13.48	-0.09	-4.88	-0.06 (0.52)
within populations	10.92	0.44	56.80	0.43 (0.00)	16.92	1.53	84.40	0.16 (0.16)
Structure 2								
source of variation	Ss	Vc	% Var	Fct/Fsc/Fst (p)	Ss	Vc	% Var	Fct/Fsc/Fst (p)
among groups	7.72	0.20	25.60	0.26 (0.00)	10.66	0.37	20.48	0.20 (0.01)
among pop within groups	8.54	0.13	16.56	0.22 (0.02)	13.48	-0.09	-4.88	-0.06 (0.52)
within populations	10.92	0.44	57.85	0.42 (0.00)	16.92	1.53	84.40	0.16 (0.16)
Structure 3								
source of variation	Ss	Vc	% Var	Fct/Fsc/Fst (p)	Ss	Vc	% Var	Fct/Fsc/Fst (p)
among groups	7.81	0.20	25.92	0.26 (0.00)	10.38	0.35	19.30	0.19 (0.00)
among pop within groups	8.45	0.12	16.15	0.22 (0.01)	13.76	-0.08	-4.43	-0.05 (0.39)
within populations	10.92	0.44	57.92	0.42 (0.00)	16.92	1.53	85.13	0.15 (0.18)
State								
source of variation	Ss	Vc	% Var	Fct/Fsc/Fst (p)	Ss	Vc	% Var	Fct/Fsc/Fst (p)
among groups	9.89	0.19	25.45	0.25 (0.00)	12.38	0.41	23.00	0.23 (0.01)
among pop within groups	6.36	0.11	14.45	0.19 (0.18)	11.76	-0.13	-7.21	-0.09 (0.58)
within populations	10.92	0.44	60.09	0.40 (0.00)	16.92	1.53	84.41	0.16 (0.16)

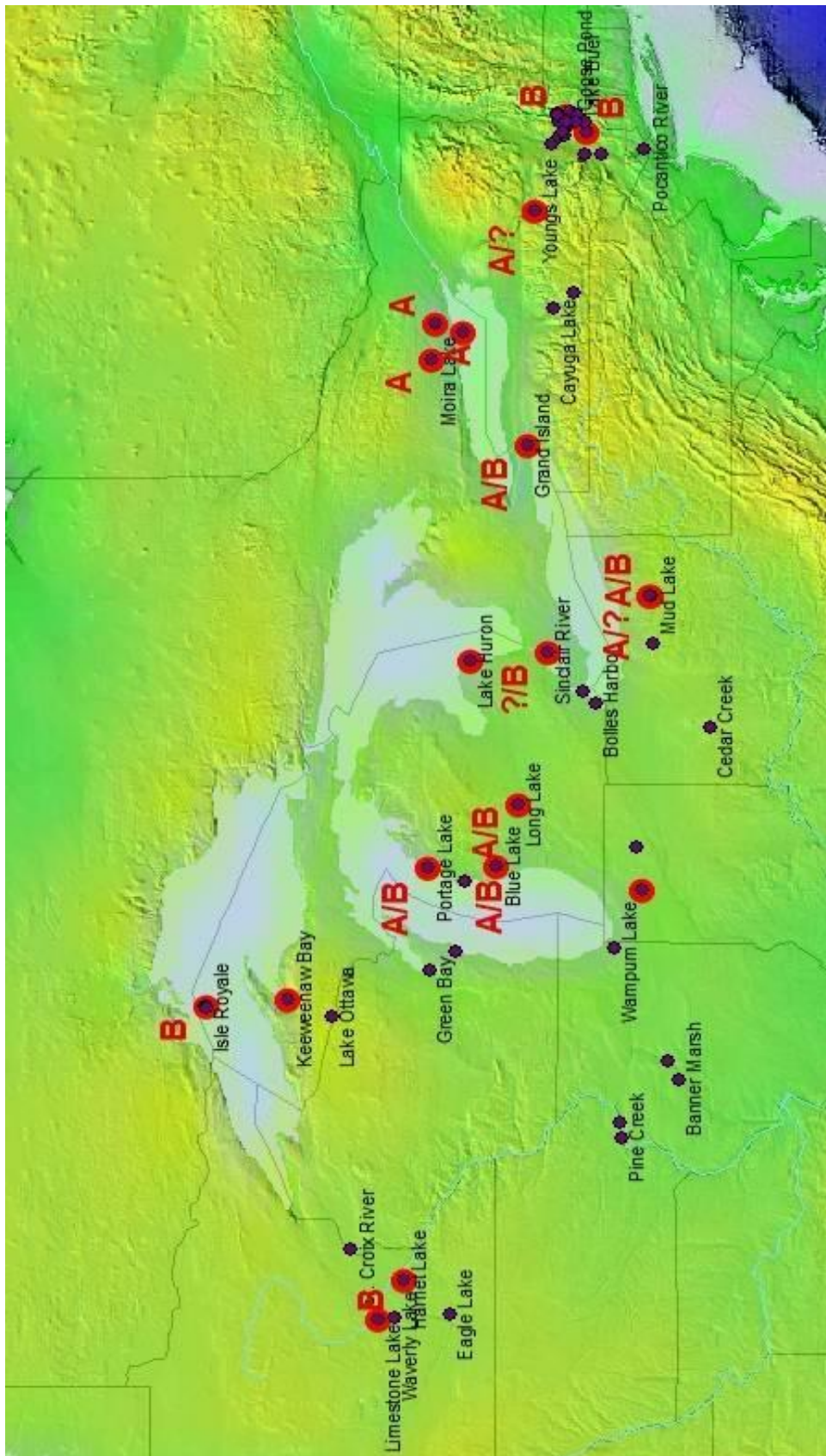
CHAPTER 5

DISCUSSION

Overall the results suggest clear structuring among the populations sampled across the range of *M. lustrica* largely independent of Euclidian distance. Less clear is the relationship between the snail, the landscape, and glaciation. To further address this question I have created two maps incorporating the phylogenetic and quantitative results, hypothetical glacial events, and actual sequence distribution across the genus' range (Figs. 20-21). In addition, I have constructed a concatenated tree with overlaid glacial time periods (Fig. 22), reinforcing the hypothetical connections between the snail's history of radiation and glacial events.

Figure 20 illustrates the broad distribution of the dominant haplotypes for COI and NDI, but also shows the limits to gene flow at the western and eastern ends of the range. COI (A) is clearly restricted to the midwest populations and Canada and is not present in the four northwest populations or in the three most eastern populations. NDI (B) is well distributed except in Canada. Both A and B are the only derived haplotypes that are widespread, indicating a common source population for those sequences. The remaining haplotypes, of which there are many, are restricted to a few populations or are singletons, suggesting rapid expansion during the retreat of the last glacier from multiple source populations.

Figure 20 (overleaf) – Map of sequenced populations and common haplotypes. Map shows the most common haplotypes for COI (A) and NDI (B). ? indicates there are no data for the respective marker. Most sample locations are shown (small dot) as well as all populations successfully sequenced (red dot), including likely other species. Red dots without a haplotype designation contain unique haplotypes, but not necessarily singletons.



The midwest populations (MI, OH, and western NY) contain multiple closely related singletons (unique haplotypes), with haplotypes A and B well represented, suggesting extended isolation of populations with subsequent remixing in multiple refugia. This pattern suggests that the Ohio Valley has acted as the central region in the radiation of the *M. lustrica*. There are three other common haplotypes for COI besides A, consisting of four sequences from Lake Superior (IRMI and KBMI), three from Harriet Lake, MN, and five sequences in western Massachusetts (SBMA and LLMA) (Appendix D). The sequences from Lake Superior and Harriet Lake are likely to be other species of *Marstonia*, while the populations in Massachusetts are closely related to haplotype A. Similarly, the two, second most common haplotypes for NDI other than B consist of the same four sequences from Lake Superior (IRMI and KBMI) and four sequences from Canada closely related to B (LOCA, MLCA, and CLCA) (Appendix J). These secondary common haplotypes, when considered in light of the distribution of the dominant haplotypes, also support regional population divisions (west, midwest, east). The northwest populations represented by multiple, distinct, and well-differentiated haplotypes (reflecting both *M. lustrica* and at least *M. lustrica* subspecies) (Figs. 12-19), further indicate that these populations may have derived from multiple source populations other than the midwest. The patterns of differentiation in Canada and the eastern New York and Massachusetts also indicate a differentiated source population from the midwest.

The observed phylogeographic pattern of the *M. lustrica* clade across its range (multiple, wide-spread singletons and few widespread haplotypes) supports multiple refugia in the wake of the most recent glacial advance. These numerous derived

singletons, in combination with the relative lack of widespread haplotypes, clearly undermines the single refugium model. The AMOVAs on watersheds and cells further support a structure indicative of multiple refugia.

In my view there are two competing hypotheses for the dispersal mechanism of *M. lustrica*: birds and pro-glacial lakes. When the statistical analysis is combined with the phylogenetic patterns the simplest explanation is the glacial lake system, with populations expanding in concert with the expansion and contraction of pro-glacial lakes. Distribution via birds should have resulted in a more panmictic pattern of dispersal, with multiple widespread common haplotypes, and the general range of *M. lustrica* should be greater, including most of New England for example.

Interpreting the phylogenetic data with a molecular clock allows for the generalized and cautious interpretation and mapping of divergent events. I have mapped all populations from this study, as well as published sequenced populations of other *Marstonia* spp. and inferred radiation pathways based on my interpretation of glacial events (Appendix G) and the use of a 1.7% molecular clock (Wilke et al 2009) (Figs. 20 & 21). Using the Tennessee Valley as a point of origin, beginning with the oldest lineages of *M. halcyon* and *agarhecta*, the biogeography and radiation of *M. lustrica* and other taxa is inferred.

The nodes on the map indicate lineage and locations of nearest known kin, and in some cases general points of origin. For example, three sequences collected during this study (LLMN, IRMI, KBMI) are associated phylogenetically with *M. comalensis*, which based on the molecular clock, derived approximately 1.0 mya. The few documented populations of this species are located in Colorado and south-central Texas (Hershler and

Liu 2011), but the specimens collected in this study would have derived from relict populations somewhere along the front of the retreating Laurentide glacier 20 kya, probably within the upper Mississippi or Ohio Valleys. The map suggests that these three representative populations radiated not through the midwest range via the Ohio Valley, but rather most likely followed the Upper Mississippi into Minnesota, accessing current Lake Superior via the Lake Duluth spillway approximately 10 kya. This radiation pathway is consistent with the timing of glacial events and the clustering of the unique haplotypes in Minnesota and Lake Superior.

Similarly, the main node representing *M. lustrica* does not represent one source population, but rather indicates that the Ohio Valley is the radiation point of origin. The map also indicates that *M. lustrica* and *M. pachyta* share a common ancestor, diverging approximately 500 kya. The time frame of 20-150 kya represents the window of opportunity that the successive advance and retreat of glaciers presented for the formation of the multiple haplotypes represented across the species current range.

The map cannot resolve the exact radiation pattern of *M. lustrica*, but suggests two main branching events, representing repeat patterns. The main range contains the numerous haplotypes in the midwest, which would have been established during the most recent retreat but are the result of successive glacial advances and retreats, and are represented on the map by the numerous short and tightly clustered red arrows. A second main branching event is represented by the red arrows into Canada and the northeast. This secondary event is consistent with the distinct haplotypes present in these populations as well as glacial history and suggests that they may be some of the most recently established populations.

In the light of glacial history, one can infer that the phylogenetic patterns observed likely reflect the effect of successive waves of glacial events. In particular the map suggests that the eastern populations likely derived via glacial Lake Maumee or Lake Warren 10-14 kya. An important note is that the *Marstonia* haplotypes collected in Minnesota and Lake Superior are likely to have been established via the upper Mississippi or through Lake Michigan. However, I believe the most likely scenario is through the upper Mississippi into Lake Superior via the Lake Duluth spillway due to the timing of glacial events and the waxing and waning of the various pro-glacial lakes, making Lake Michigan an unlikely pathway.

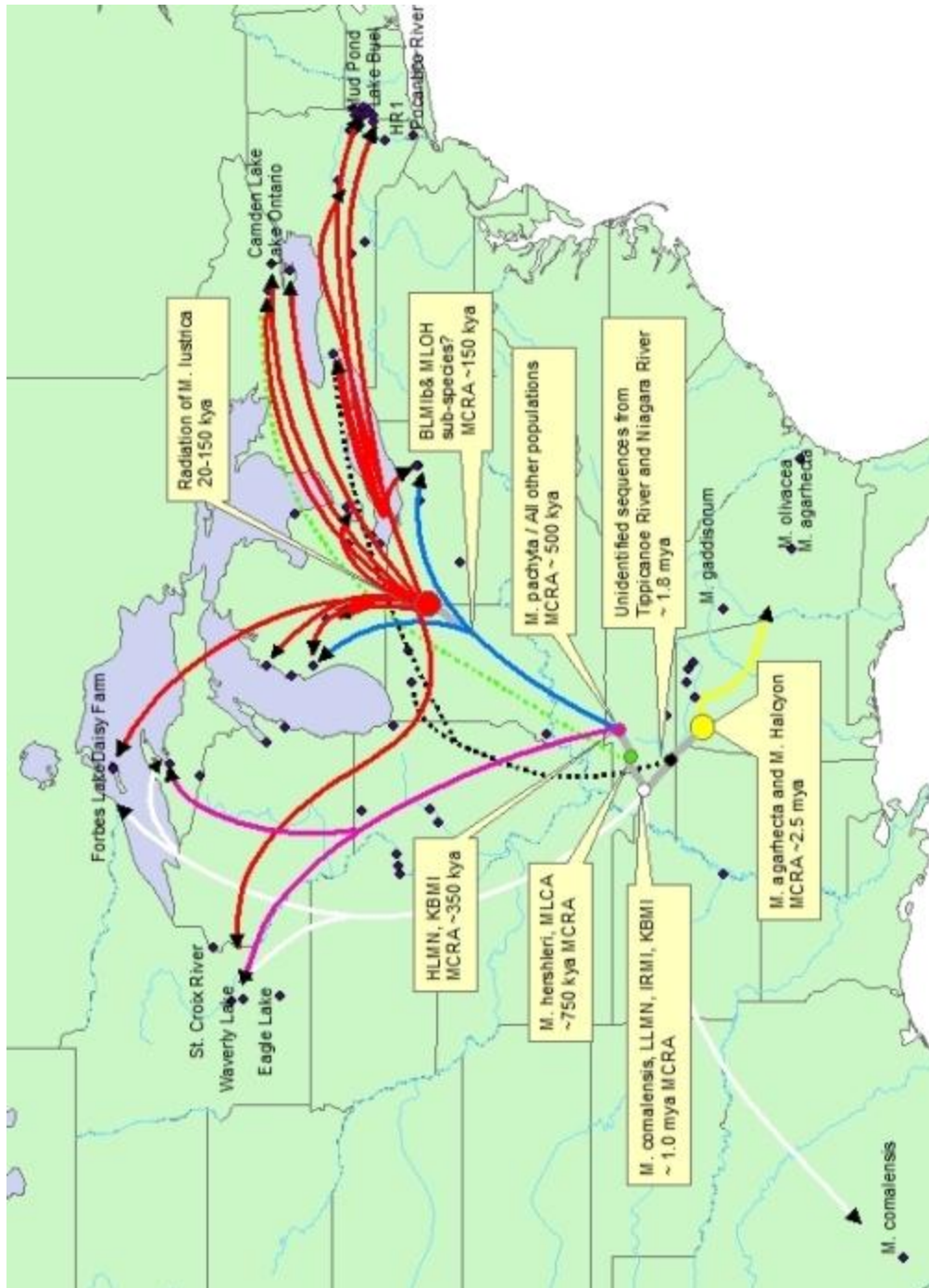


Figure 21 – Map of hypothesized species relationships and divergent events. Map shows locations of populations with hypothetical lineages based on phylogenetic analysis with a molecular clock of 1.7%/myr.

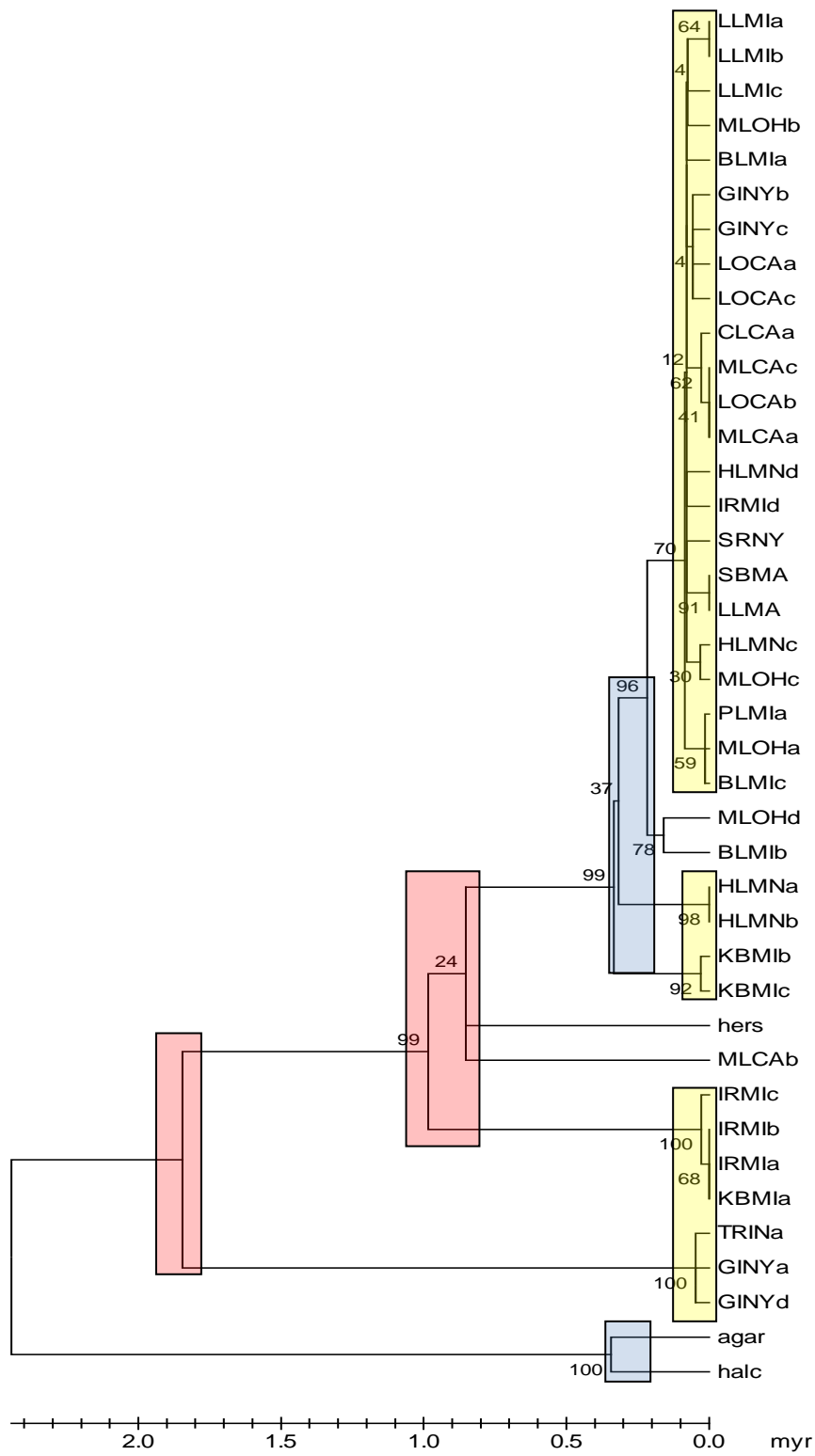


Figure 22 – Phylogenetic tree with associated glacial time periods (yellow = Sangamon interglacial; blue = Illinoian stage; red = pre-Illinoian stage).

CHAPTER 6

CONCLUSION

At the macro level, the null model stated that Euclidian geographic distance would be the best predictor for population structuring, reflecting an interpretation of glacial history whereby all current populations of *M. lustrica* represent one radiation event from a geographically isolated source population. The alternative model suggests that the current distribution of *M. lustrica* is the result of multiple glacial advances and retreats, multiple source populations, and contemporary isolation.

The Mantel tests on Fst and Euclidian distance did not strongly relate to COI or NDI genetic structure (Tables 9 & 10), but were significant. This is consistent with the notable haplotype diversity and distribution pattern for both markers, and supports a glacial processes of dispersal. The watershed model is a better explanation for structure than Euclidian distance alone, and is borne out by the AMOVA results (Table 13), demonstrated by the relatively strong performance of the watershed and cell-level analyses. While cell structure performed better than watersheds, the watershed structure is inherent in the cell structure, while cells reflect smaller clusters of populations with lower within group variability. The ambiguity of the cell structure is also more amendable to the hypothesized and therefore uncertain boundaries of glacial lakes, whereas the USGS watershed structure is based on contemporary watershed boundaries.

The diversity of haplotypes observed does not support a single radiation event, rather the multiple lineages and patterns represented across the range of *M. lustrica* clearly indicate multiple advances and retreats, out of and into multiple source

populations. Such a structure can only be explained by multiple divergent events over very long time frames.

At the meso scale, the null model stated that the observed genetic structure of the eastern populations should reflect panmixia, resulting from the recent mixing of populations due to human and non-human “bucket” events. Collectively, these populations would also reflect the null model of Euclidian geographic distance relative to the remaining populations across the species range, assuming that any bucket events would have been regionally based. The alternative model stated that the eastern populations collectively are more closely related to each other than to the midwest or western populations, and that they would reflect not panmixia but rather unique haplotypes not represented in other regions due to long-term isolation and glacial events.

Clearly panmixia is not occurring in the eastern region. The Mantel test on distance for the eastern populations was stronger than for the combined data set (Table 9), suggesting that distance was a factor in the genetic structure of those populations, but these results need to be viewed with caution due to the small dataset. Specifically, the significant distance relationship is likely due to the lack of haplotype diversity in the NY and MA populations relative to the Canadian populations. This difference is not insignificant: population differentiation in the east is not consistent with the rest of the range and all four populations (YLNy, SRNY, SBMA, LLMA) consisting of relatively large sample sizes contained only one haplotype each.

There is also no question that the two populations in Massachusetts and the SRNY population are more closely related to each other relative to the populations in western NY or Canada, or to any other population. It is clear that each of the three trees

(Figs. 12-14) cluster these populations, albeit slightly differently in each case. The concatenated tree clusters the Canadian populations separately from the LLMA, SBMA and SRNY populations, supporting sub-regional differentiation within the eastern range. However, that tree also includes one sequence each from HLMN and IRMI, closely related to the SRNY and MA populations, making these five populations more closely related to each other than their regional counterparts. The NDI tree results in a similar divergence, with HLMN being clustered with SRNY and the MA populations, while clustering the Canadian populations at the ends of the clade. The COI tree also clusters the SRNY and MA populations, but does so by putting them at separate ends of the clade. In short, the results are confounded, but suggest some form of sorting among the eastern populations, and clearly suggest a recent common ancestor for the western and eastern populations.

The phylogenetic trees, networks, and quantitative analysis support strong geographical structuring of the populations. The most obvious is the high number of older taxa and the clear delineation of haplotypes in the northwest (MN & Lake Superior) and the rest of the range. Similarly, the eastern populations (YLN, SRNY, and MA) clearly demonstrate the lowest haplotype diversity (COI) of all the populations represented by more than one sequence, all containing one unique haplotype and only being separated by a single bp. The implication is that this region is the most recent to be inhabited by *M. lustrica*, possibly from a different refugium. However, it is also possible that the lack of a common haplotype for the eastern and midwest populations may have been either lost to drift, or may have been missed during sampling. Quantitatively, the AMOVAs show the

strongest structuring takes place at the watershed and cell levels, both of which reflect natural features of isolation, and are consistent with effects of glaciation.

Simultaneously, the results clearly undermine recent gene flow, specifically demonstrating that the numerous unique haplotypes and singletons are fully consistent with much older processes of fixation and isolation since at least the Illinoian glacier. Such an outcome clearly indicates that random transport by agents other than glacial lakes is very rare, and thoroughly undermines the role of humans as agents of recent dispersal. In particular, it is reasonable to assume that given the data here, the populations in eastern New York and Massachusetts are unique and were either established before the arrival of Europeans, or are from an unidentified source population.

Under the null model it was anticipated that given the higher density of historical populations in the midwest, in combination with the common suggestion that gastropods are distributed via birds, that those populations would exhibit little differentiation. Other than a general grouping at the watershed level, there does not appear to be any structure resembling panmixia, which is not what one would expect if there was active gene exchange occurring between populations. These results are consistent with the alternative models suggesting that population structure is the result of glacial processes establishing populations, and that watersheds continue to play a role in isolation.

Using COI, Hershler et al (2003) clearly articulated the phylogeny of seven *Marstonia* (*M. lustrica*, *M. pachyta*, *M. hershleri*, *M. comalensis*, *M. halcyon*, *M. castor*, *M. agarhecta*) to other nymphophiline. Here that work has been built upon, clearly showing a tight, yet unresolved clade for *M. lustrica*, while adding significant structure to the genus as a whole, and shedding light on its relationship to glacial processes. By

sequencing multiple populations across the range of *M. lustrica* using COI and NDI, a robust phylogenetic pattern has been developed, while simultaneously identifying populations whose status is now in question. Clearly the phylogeography of *Marstonia* is more complicated than previously thought.

Retreating glaciers resulted in significant pro-glacial lake development, providing opportunities for expansion of aquatic organisms on numerous occasions and over vast areas. As a result, contemporary populations at the southern edge of *M. lustrica*'s range represent relatively recent introductions dating from about 150 kya, with more northerly populations representing introductions within 10-20 kya, most notable perhaps being the eastern populations in Canada, New York and Massachusetts.

The presence of numerous unique haplotypes (singletons) across the range of *Marstonia*, differing in most cases by only a few bp, indicate rapid recent expansion associated with post glacial expansion (Emerson and Hewitt 2005). In the case of *M. lustrica* the process of divergence takes place during the interglacials, with the rapid retreat of glaciers providing dispersal opportunities in pro-glacial lakes and subsequent isolation of the snail, typically lasting for tens of thousands of years. Temperature shifts of 7-15 C can take place in as little as a few decades, but last for thousands of years, providing dramatic changes in the distribution of glaciers (Hewitt 2004), and resulting in fast distribution, subsequent isolation, and ultimately the pruning of haplotypes.

Specifically, the Sangamon interglacial (150 kya), would have provided the most recent genetically recordable episode of large-scale expansion and divergence, allowing for multiple haplotypes (of 1-2 bp) to become fixed within the various populations. With the onset of the Wisconsinan glacier 80-100 kya, reaching its maximum extent on several

occasions, these dispersed populations would have been trimmed and remixed along the front of the advancing glacier. Following the retreat of the Wisconsinan, these mixed populations would have again undergone dispersal and subsequent isolation based on the retreat patterns of the glacier. The most common derived haplotypes, and all the haplotypes with only 1-2 bp difference, arose within the time frame of the Sangamon interglacial and the maximum extent of the Wisconsinan, and reflect their historical origins across the front of the retreating glacier.

Based on these assumptions of glacial history, combined with the phylogenetic trees, networks, and quantitative analysis described here, it appears there are three main groupings of *M. lustrica*; the northwest, midwest, and the eastern populations (Canada, eastern New York and western Massachusetts). Possible associated refugia are the upper Mississippi, the Ohio valley, and the eastern seaboard (possibly Lake Connecticut or a more southern system). It is clear that *M. lustrica* originally evolved alongside *M. pachyta* approximately 0.5 mya, within a glacial refugium in the Tennessee and Ohio Valleys, and that all of the other populations have radiated out of the Ohio Valley refugium since that time.

The remaining haplotypes found outside of the main clade and across the range likely represent other species of *Marstonia*, and likewise represent historical mixings at southern contact zones. Because these areas of contact would by definition have been at the fringes of each species' ranges, their inclusion in pro-glacial lake systems and overlap with the range of *M. lustrica* would have been rare, which is consistent with results here.

In conclusion, the high number of unique haplotypes, with only one common derived haplotype for COI and NDI each, suggests that *M. lustrica* has not undergone

contemporary gene flow, but that gene flow has been restricted both regionally and within watersheds. The dominant haplotypes that are well distributed are indicative of a pro-glacial event that was mostly restricted to the core range, and most likely emerged out of the Ohio valley during the early Wisconsinan.

Marstonia spp. exhibit greater diversity in the south, suggesting stability and isolation of those populations over hundreds of thousands of years. Unlike the southern species, *M. lustrica* exhibits shallow divergence across a much greater geographic area, strongly supporting a recent expansion and the likely role of glaciation in its evolution. This work has served to clarify standing questions about the landscape ecology and biogeography of *M. lustrica*, while raising new questions about the biogeographical relationship between the snail and its sister species.

Conservation

Given the endangered status of many gastropods, including several of the *Marstonia* spp. discussed here, it seems reasonable to pay greater attention to the condition and status of *M. lustrica*. This is particularly prudent given the phylgeographic results reported here: with multiple species across the range and clear genetic structuring within *M. lustrica*, there is much yet to be learned about the *Marstonia* genus as whole. In addition, given *M. lustrica*'s association with cold water systems, the clear trend of globally rising temperatures may be a significant risk factor for the species.

In light of these results, it seems rational to develop a conservation strategy that recognizes the genetic variability represented across the range within *M. lustrica* as well as for the whole genus. Focusing solely on *M. lustrica*, one option is to designate

regionally based conservation management zones, with an eye towards cataloging, monitoring, and maintaining as much genetic diversity as possible. Designating the northwestern, midwestern, and eastern populations as distinct conservation zones would reflect both their genetic differentiation as well as our understanding of the likely continued isolation of those regional populations from each other. A more holistic approach may be to designate the whole genus as a conservation unit. Such a plan would recognize 1) that while *M. lustrica* as a species is well distributed, it does exhibit geographic structuring as articulated above, and 2) the genus as a whole is very rare (with many individual species already listed at the federal level) and that the evolutionary processes driving their rarity is likely to be at play for each of them despite our lack of understanding.

Not only is this relevant for the conservation of *Marstonia*, but I believe it is relevant to helping us further understand the role of glaciation in the development of watersheds and aquatic systems, with attendant implications for speciation of numerous other obligate aquatic organisms.

Recognizing the shortcomings of this project, there are a number of things that need to be pursued to deepen our understanding of the phylogeography of the *Marstonia* genus. Ideally we need to complete a more robust sampling of the Great Lakes associated waterbodies to more fully articulate the phylogenetics of *M. lustrica* and to determine the true extent of other species across its range. This can be done using mtDNA markers but should be accompanied by additional markers including RNA and/or nuclear DNA. Such an analysis would provide greater resolution at the population level and allow us to infer a more detailed phylogeography.

In lieu of a broad scale analysis, I think it would be prudent to conduct several intensive studies on those few waterbodies that exhibited multiple haplotypes and species in this study, using the additional markers. This scale of work could be coupled with detailed anatomical analysis to help clarify our conceptualization of *Marstonia* species, and may also help to elucidate any adaptive traits for the purpose of reinforcing a given conservation strategy.

At a more basic level, it would be ideal to conduct some long term studies on captive populations in an effort to better understand the basic biology and life cycle of *Marstonia*. Such work could include experiments with temperature regimes in an effort to place the conservation of the genus in the context global climate change.

APPENDIX A

LIST OF LAKES, PONDS, AND STREAMS SAMPLED IN 2000 AND 2005.

New York:

- Barrett Pond
- Flatbrook
- Kinderhook Lake
- Young Lakes (Weaver Lake & Young Lake, type locality)
- Pocantico River
- Queechy Lake
- Saw Kill (known location)
- Shaker Swamp
- Smith Pond
- Sutherland Pond

Massachusetts:

- Card Pond
- Cranberry Pond
- Crane Lake
- Goose Pond
- Lake Buel
- Lake Mansfield
- Laurel Lake
- Mud Pond
- Onota Lake
- Pontoosuc Pond
- Richmond Pond
- Shaker Mill Pond
- Stockbridge Bowl (known location)

Connecticut:

- Twin Lakes

APPENDIX B

WATER BODIES SAMPLED IN 2007 AND 2008.

Catatonk Creek, NY	Waverly Lake, MN
Cayuga Lake, NY	Forbes Lake, MI
Hudson River (18 sites), NY	Lake Ojibway, MI
Niagara River, NY	Daisy Farm, MI
Pocantico River, NY	Lake Benson, MI
Mud Lake, OH	Keeweenaw Bay, MI
Long Lake, OH	Lake Ottawa (Pickerel Lake), MI
Cedar Creek, OH	Lake Michigan, Ludington, MI
Bailey Lakes, OH	Blue Lake, MI
Banner Marsh (Johnson Lake), IL	Saint Clair River, MI
Illinois River, Rock Island, IL	Bolles Harbor, MI
Mississippi River, IL	Lake Erie Metropark, MI
Wampum Lake, IL	Long Lake, MI
Tippicanoe River, IN	Lake Huron (Harbor Beach area), MI
Winona Lake, IN	Portage Lake, MI
Pine Creek, IO	Lake Michigan, Manitowoc, WI
Eagle Lake, MN	Greenbay, WI
Harriet Lake, MN	Camden Lake, CN
St. Croix River, MN	Lake Ontario, CN
Limestone Lake, MN	Moir Lake, CN

APPENDIX C

COI SEQUENCE TEXT FILES.

```
#mega !Title : COI.TXT;
!Format DataType=DNA indel=- CodeTable=Standard;

!Domain=Data property=Coding CodonStart=1;
#HLMNa
CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCACTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTTCCTCGTGATACCGATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTCCTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCACCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTAAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#HLMNb
CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCACTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTTCCTCGTGATACCGATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTCCTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCACCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTAAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#HLMNc
CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTTCCTAGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAG-----

#HLMNd
-----GGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTTCCTAGTGATACCAATAATAATTGG
GGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#HLMNe
-----GGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCACTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTTCCTCGTGATACCGATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTCCTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCACCACTCTCTAGAAATCT
```

TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTAAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTAGCCG-----

#IRMla

CACTTTATACATTCTATTTGGCATATGATCTGGCCTAGTCGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGACAACAGGCGCTCTCCTAGGAGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTCTCGTGATACCGATAATAATTGG
TGGTTTTGGCAACTGGCTTGTACCTTTGATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTCCTGCTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTGCTAGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGAGGCGACCCAATTTTATACCAGCACTTATTT

#IRMlb

CACTTTATACATTCTATTTGGCATATGATCTGGCCTAGTCGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGACAACAGGCGCTCTCCTAGGAGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTCTCGTGATACCGATAATAATTGG
TGGTTTTGGCAACTGGCTTGTACCTTTGATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTCCTGCTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTGCTAGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGAGGCGACCCAATTTTATACCAGCACTTATTT

#IRMlc

-----GGCCTAGTCGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGACAACAGGCGCTCTCCTAGGAGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTCTCGTGATACCGATAATAATTGG
TGGTTTTGGCAACTGGCTTGTACCTTTGATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTCCTGCTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTGCTAGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGAGGCGACCCAATTTTATACCAGCACTTATTT

#IRMld

-----CTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTCTCGTGATACCAATAATAATTGG
GGGTTTTGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAGTCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGCGGAGGCGACCCAATTTTATATCAGCACTTATTT

#GINYa

-----TAGTTGGAACAGCTTTAAGTTTATT
AATTCGAGCAGAATTAGGGCAACAGGCGCTCTTCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGGCCACGCTTTTCGTCATAATTTTTTCTAGTAATACCAATAATAATTGG
AGGTTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGGACCAGACATGGCATTCCCCCG
ACTAAACAATATAAGATTTTGACTTTTACCTCCTGCCCTATTACTCCTGCTATCCTCAGC
TGCAGTAGAAAGAGGCGTTGGGACTGGATGAACCGTTTATCCACCCTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTTGACTTAGCCATTTTTTCCCTGCACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAMCAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACGGCTATTTTGTTATT
GTTATC-----

#GINYb

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTCATAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#GINYc

-----TATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTCATAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#GINYd

-----GCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTATT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTCATAATTTTTTTCCTAGTAATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAACAATATAAGATTTTGACTTTTACCTCCTGCCCTATTACTCCTGCTATCCTCAGC
TGCAGTAGAAAAGAGGCGTTGGGACTGGATGAACCGTTTATCCACCAGTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTTGACTTAGCCATTTTTTCCCTGCCTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTATT
GTTATCTCTACCAAGTGTAGCCGGTGCCATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTTGACCCGGCGGGCGGGGGGAGACCCAATTTTATACCAGCACTTGT--

#MLOHa

-----ATTGTCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTCATAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCA-----

#MLOHb

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTCATAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#MLOHc

-----TTGGAACAGCTTTAAGCTTGCT

AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGCACGCTTTTCGTCATAAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCACTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#MLOHd

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGCACGCTTTTCGTCATAAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGACCAGACATGGCATTCCCCCG
GCTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCACTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#LLMIa

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGCACGCTTTTCGTCATAAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCACTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#LLMIb

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGCACGCTTTTCGTCATAAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCACTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#LLMIc

-----GCCTAGTTGGAACAGCTTTAAGCTTGTT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGCACGCTTTTCGTCATAAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCACTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCGCTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTGATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCA-----

#LOCAa

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTGCACGCTTTTCGTCATAAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGACCAGACATGGCATTCCACG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC

TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTCCTGATACCAATAATAATTGG
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTCGACCCGGCGGGTGGGGGCGACCCAATTTTATATCAGCACTTATTT

#LOCAb

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAC TGCCACGCTTTTCGTCATAAATTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTCCTGATACCAATAATAATTGG
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTCGACCCGGCGGGGCGGGGCGACCCAATTTTATACCAGCACTTATTT

#LOCAc

-----CCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAC TGCCACGCTTTTCGTCATAAATTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCACG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTCCTGATACCAATAATAATTGG
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTCGACCCGGCGGGGCGGGGCGACCCAATTTTATATCA-----

#MLCAa

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAC TGCCACGCTTTTCGTCATAAATTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTCCTGATACCAATAATAATTGG
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTCGACCCGGCGGGGCGGGGCGACCCAATTTTATATCAGCACTTATTT

#MLCAb

CACTTTATATATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTTCGAGCAGAATTAGGGCAGCCAGGCACTCTCTAGGAGACGACCAGCTATACAACGT
AATTGTAAC TGCCACGCTTTTCGTCATAAATTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGTAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTCTGACTTTTGCCCCCTGCCCTATTACTCTGTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGGTGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTCCTGATACCAATAATAATTGG
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGGCGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTAAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTCGACCCGGCGGGGCGAGGCGATCCAATTTTATACCAGCACTTATTT

#MLCAc

-----TTGGAACAGCTTTAAGCTTGCT
AATTTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAC TGCCACGCTTTTCGTCATAAATTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTCCTGATACCAATAATAATTGG
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT

ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#BLMIa

-----GGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCACTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCA-----

#BLMIb

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCACTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTCTAATACTTGGAGCACCAGACATGGCATTCCCCCG
GCTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCATTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTAT--

#BLMIc

-----TTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCACTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGCTGGGACTGGATGAACGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#BLMId

-----TTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCACTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#KBMIa

CACTTTATACATTCTATTTGGCATATGATCTGGCCTAGTCGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGACAACCAGGCGCTCTCCTAGGAGACGACCACTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCGATAATAATTGG
TGGTTTTCGGCAACTGGCTTGTACCTTTGATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTCTGCTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTTCCCTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGGCGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTGCTAGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGAGGCGACCCAATTTTATACCAGCACTTATTT

#KBMIb

-----TGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAAATTTTTTCTCGTGATACCGATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTCCTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCA-----

#KBMIc

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAAATTTTTTCTCGTGATACCGATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTCCTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCA-----

#CLCAa

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAAATTTTTTCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#CLCAb

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAAATTTTTTCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#CLCAc

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAAATTTTTTCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCCTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#LLOHa

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAAATTTTTTCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG

ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGCGACCCAATTTTATATCAGCACTTATTT

#LLOHb

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAATTTTTCCTCGTAATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
GCTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGCCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGCGACCCAATTTTATATCAGCACTTATTT

#LLOHc

-----AAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAATTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTT-----

#LLMNa

-----CGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGAGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAATTTTTCCTCGTGATACCGATAATAATTGG
AGGTTTGGCAATTGGCTTGTACCTTTGATACTTGGGGCACCAGACATGGCATTCCCTCG
ACTAAATAATATAAGATTTTGACTTTTACCTCCTGCCCTATTACTCCTGCTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGGCGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTAAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTGTCTAGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGAGGCGACCCAATTTTATACCACTACT-----

#LLMNb

-----CTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGAGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAATTTTTCCTCGTGATACCGATAATAATTGG
AGGTTTGGCAATTGGCTTGTACCTTTGATACTTGGGGCACCAGACATGGCATTCCCTCG
ACTAAATAATATAAGATTTTGACTTTTACCTCCTGCCCTATTACTCCTGCTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGGCGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTAAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTGTCTAGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGAGGCGACCCAATTTTATACCACTACTTTT

#PLMIa

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTCGTCATAATTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT

GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGC-----

#TRINa

-----TAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTACATAATTTTTTCCTAGTAATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTGGGGGCACCAGACATGGCATTTCCTCCG
ACTAAACAATATAAGATTTTGACTTTTACCTCCTGCCCTATTACTCCTGCTATCCTCAGC
TGCAGTAGAAAAGAGGCGTTGGGACTGGATGAACCGTTTATCCACCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTTGACTTAGCCATTTTTTCCTGCACCTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGTTATT
GTTATCTCTACCACTGCTAGCCGGTGCCATTACAATACTTTTAACCGATCGAAAATTTAA
TACCGCGTTTTTGACCCGGCGGGCGGGGAGACCCAATTTTATACCAGCACTTGTTT

#SBMAa

-----GCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTACATAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTGGGGGCACCAGACATGGCATTTCCTCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGGAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGCGGACCCAATTTTATACAGCACTTATTT

#SBMAb

-----GCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTACATAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTGGGGGCACCAGACATGGCATTTCCTCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGGAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGCGGACCCAATTTTATACAGCACTTATTT

#SBMAc

-----GGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTACATAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTGGGGGCACCAGACATGGCATTTCCTCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGGAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGCGGACCCAATTTTATACAGCACTTATTT

#M.lustrica2000 -----AGTTGGAACAGCTTTAAGCTTGCT

AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTACATAATTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTGGGGGCACCAGACATGGCATTTCCTCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGGAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGCGGACCCA-----

#YLNya

-----AAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#YLNyB

-----AAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTA-----

#YLNyC

-----AAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTA-----

#YLNyD

-----AAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#LLMAa

-----TTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGGAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTTACACTTAGCCGGGTGTCTC
CTCAATTTTAGGGGCAGTAAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAAATTTTAA
TACCGCGTTTTTCGACCCGGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#LLMAb

-----ACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAACTGCCCACGCTTTTCGTCATAAATTTTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG

ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCGCCACTCTCTAGGAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGACCGCGGGCGGGGGCGACCCAATTTTATATCAGCACTTATTT

#SRNYa

----CGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTACATAATTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCACCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATC-----

#SRNYb

CACTTTATACATTCTATTTGGCATATGATCCGGCCTAGTTGGAACAGCTTTAAGCTTGCT
AATTCGAGCAGAATTAGGGCAACCAGGCGCTCTCCTAGGGGACGACCAGCTATACAACGT
AATTGTAAGTACGACGCTTTTCGTACATAATTTTTCCTCGTGATACCAATAATAATTGG
AGGTTTCGGCAATTGGCTTGTACCTTTAATACTTGGGGCACCAGACATGGCATTCCCCCG
ACTAAATAATATAAGATTTTGACTTTTGCCCCCTGCCTTATTACTACTACTATCCTCAGC
TGCAGTAGAAAGAGGAGTTGGGACTGGATGAACCGTTTATCCACCACTCTCTAGAAATCT
TGCTCATGCCGGCGGGCTCAGTCGACTTAGCCATTTTTCCTACACTTAGCCGGTGTCTC
CTCAATTTTAGGGGCAGTAACTTTATTACAACAATCATCAATATGCGATGACGGGGCAT
GCAATTTGAACGATTGCCTTTATTCGTATGATCTGTGAAAATTACTGCTATTTGCTTTT
ATTATCTCTGCCAGTACTGGCCGGTGCAATTACAATACTTTTAACCGATCGAAATTTTAA
TACCGCGTTTTTCGA-----

APPENDIX D

NDI SEQUENCE TEXT FILES.

>HLMNa

TTTTTACCCTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAAACCAACAATGGCTAATTTTTCTCCTTA
TTTTTGGGCCCCAGTATTACAGCTTTATTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTT
TAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATATGGCACACTTTTAGCAGGATGAGCATCAAAC
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTTTCATTTATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>HLMNb

TTTTTACCCTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAAACCAACAATGGCTAATTTTTCTCCTTA
TTTTTGGGCCCCAGTATTACAGCTTTATTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTT
TAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATATGGCACACTTTTAGCAGGATGAGCATCAAAC
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTT

>HLMNc

TTTTTACTTTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGCT
ACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTAT
TTTTTGGGCCCCAGTATTACAGCTTTATTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAAC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAA
TACTTCCGCTTTCATTCATATGATTTGTAACCTGTATTGCCGAAACCAATCGA

>HLMNd

CCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTATTTTTGGGCCCCAGTATTACAGC
TTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTTAAATGAGGCATCTTATTCTTT
CTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCTCTAAATATGCCTTACTAGGAA
GCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTCTTCCCTTTATTTATTATGACC
ACATTCAGATACATTGAACTAAATGAAACCAAGAAGCAGTGTGGATAACATTTCTAATACTTCCGCTTTCATTCATAT
GATTTGTAACCTGTATCGCCGAAACCAATCGA

>IRMIa

TTTTTACCCTATTAGAACGAAAGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGTCTTGCAGGGCT
GCCTCAGCCCCCTAGCAGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAGCCAACAATGGCTAATTTTTCTCCTTAT
TTTTTGGGCCCCAGTATTACAGTTTATTTTAGCCCTCTTATTATGACAATTATACCTAGACTTTTCTCCACCAGATACTTT
AAATGGGGCATCTTATTCTTTCTTTGTGTTTCAAGATTAAACGTATATGGTACACTTTTAGCAGGATGAGCATCAAAC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATAACCACATTCAGATACATTGAACTAAATGAAAGTCAAGAAGCAGTATGGATAACATTTCTAA
TGCTCCCGCTTTCATTCATGTGGTTTGTAACTGTATCGCCGAAACCAATCGG

>IRMIb

TTTTTACCCTATTAGAACGAAAGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGTCTTGCAGGGCT
GCCTCAGCCCCCTAGCAGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAGCCAACAATGGCTAATTTTTCTCCTTAT
TTTTGGGCCCCAGTATTACAGTTTATTTTAGCCCTCTTATTATGACAATTATACCTAGACTTTTCTCCACCAGATACTTT
AAATGGGGCATCTTATTCTTTCTTTGTGTTTCAAGATTAAACGTATATGGTACACTTTTAGCAGGATGAGCATCAAAC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATAACCACATTCAGATACATTGAACTAAATGAAAGTCAAGAAGCAGTATGGATAACATTTCTAA
TGCTCCCGCTTTCATTCATGTGGTTTGTAACTGTATCGCCGAAACCAATCGG

>IRMIc

TTTTTACCCTATTAGAACGAAAGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGTCTTGCAGGGCT
GCCTCAGCCCCCTAGCAGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAGCCAACAATGGCTAATTTTTCTCCTTAT
TTTTGGGCCCCAGTATTACAGTTTATTTTAGCCCTCTTATTATGACAATTATACCTAGACTTTTCTCCACCAGATACTTT
AAATGGGGCATCTTATTCTTTCTTTGTGTTTCAAGATTAAACGTATATGGTACACTTTTAGCAGGATGAGCATCAAAC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATAACCACATTCAGATACATTGAACTAAATGAAAGTCAAGAAGCAGTATGGATAACATTTCTAA
TGCTCCCGCTTTCATTCATGTGATTTGTAACCTGTATCGCCGAAACCAATCGG

>IRMId

TTTTTACCCTATTAGAACGAAAGGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCAGTATTAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTT
TAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAAC
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
TACTTCCGCTTTCATTTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>GINYa

TTTTTACCTTATTAGAACGAAAGGGGGCTATCTTATATGCAGATTTCGTAAAGGACCAAATAAAAGTAGGCCTTGCAGGGCT
ACCTCAGCCCCCTAGCAGATGCCGCTAAGCTTCTTACCAAAGAAATCGCAAAACCAACAATGGCTAATTTTTCTCCTTAT
TTTTGAGCCCCAGTATTAGTTTATTTTAGCCCTTTTATTATGACAATTATACCCAGACTTTTCTCCACCGGGTACTTT
AAATGGGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATATGGTACACTTTTAGCAGGATGGGCATCAAAC
TAAATATGCCTTATTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATCAGCATAGCACTTATTCTCCTT
TTTCCCTATTATTATAACCACATTCAGATACATTGAACTAAATGAAAGACAAGAAGCAGTGTGAATAACATTTCTAA
TGCTTCCACTTTCATTTATATGATTTGTAACCTGTATTGCCGAGACCAACCGA

>GINYb

CTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGCTACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAG
AAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTATTTTGGGCCCCAGTATTAGCTTTATTTTAGCCCTCTTATTAT
GACAATTATACCCAGACTTTTCTCCACCAGATACTTTAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAAC
GTATACGGGCACACTTTTAGCAGGATGAGCATCAAACCTCTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACA
ACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTCTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACT
AAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAATACTTCCGCTTTCATTTCATATGATTTGTAACCTGTATCGCC
GAAACCAATCGA

>GINYc

TTTTTACCCTATTAGAACGAAAGGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCAGTATTAGCTTTATTTTAGCCCTTTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTT
TAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAAC
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
TACTTCCGCTTTCATTTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>GINYd

TTTTTACCTTATTAGAACGAAAGGGGGCTATCTTATATGCAGATTTCGTAAAGGACCAAATAAAAGTAGGCCTTGCAGGGCT
ACCTCAGCCCCCTAGCAGATGCCGCTAAGCTTCTTACCAAAGAAATCGCAAAACCAACAATGGCTAATTTTTCTCCTTAT
TTTTGAGCCCCAGTATTAGTTTATTTTAGCCCTTTTATTATGACAATTATACCCAGACTTTTCTCCACCGGGTACTTT
AAATGGGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATATGGTACACTTTTAGCAGGATGGGCATCAAAC
TAAATATGCCTTATTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATCAGCATAGCACTTATTCTCCTT
TTTCCCTATTATTATAACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGAATAACATTTCTAA
TGCTTCCACTTTCATTTATATGATTTGTAACCTGTATTGCCGAGACCAACCGA

>MLOHa

TTTTTACTTTATTAGAACGAAAGGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGCT
ACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTAT
TTTTGGGCCCCAGTATTAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAAC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGAATAACATTTCTAA
TACTTCCGCTTTCATTTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>MLOHb

TTTTTACCCTATTAGAACGAAAGGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCAGTATTAGCTTTATTTTAGCCCTTTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTT
TAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAAC
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
TACTTCCGCTTTCATTTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>MLOHc

TTTTTACTTTATTAGAACGAAAGGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGCT
ACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTAT
TTTTGGGCCCCAGTATTAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAAC

TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTG

>MLOHd

TTTTTACCCTATTAGAACGAAAGGGGCTATCTTATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGCT
ACCCAGCCCCCTAGCAGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAAACCAACAATGGCTAATTTTTCTCCTTAT
TTTTGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
TAAATATGCTTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTG

>LLMIa

TTTTTACCCTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
TAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATTTCTTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>LLMIb

TATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGCTACCCAGCCCCCTAGCAGATGCCGCTAAACTT
CTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTATTTTTGGGCCCCAGTATTCAGCTTTATTTTAGC
CCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTTAAATGAGGCATCTTATTCTTTCTTTGTGTTTC
AAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCTCTAAATATGCCTTACTAGGAAGCCTACGGGC
CATCGCACAAACAATTTCTTATGAAATTAGCATAGCACTTATCTTACTCTTCCCTTTATTTATTATGACCACATTCAGAT
ACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAATACTTCCGCTTTCATTCATATGATTTGTAAC
CTGTATCGCCGAAACCAATCGA

>LLMIc

TTTTTACCCTATTAGAACGAAAGGGGCTATCGTATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
TAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>LOCAa

TTTTTACCCTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
TAAATGAAAGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATATATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>LOCAb

TTTTTACCCTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCGGATACTTT
TAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>LOCAc

GTAGGAATTGCAGGGCTACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATG
GCCAATTTTTCTCCTTATTTTTGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTT
TCTCCACCAGATACTTTAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGC
AGGATGAGCATCAAACCTCTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAG
CATGCACTTATCTTACTCTTCCCTTTATTTATTATGACCACATTCAGATATATTGAACTAAATGAAAGCCAAGAAGCA
TGTGGATAACATTTCTAATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>MLCAa

TATATGCAAACCTCGTAAAGGACCAAATAAAAGTAGGAATTGCAGGGCTACCCAGCCCCCTAGCAGATGCCGCTAAACTT
CTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTATTTTTGGGCCCCAGTATTCAGCTTTATTTTAGC
CCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCGGATACTTTAAATGAGGCATCTTATTCTTTCTTTGTGTTTC
AAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCTCTAAATATGCCTTACTAGGAAGCCTACGGGC

CATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTCTTCCCTTTATTTATTATGACCACATTCAGAT
ACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAATACTTCCGCTTTCATTCATATGATTTGTAAC
CTGTATCGCCGAAACCAATCGA

>MLCAb

TTTTTACCTTATTAGAACGAAAGGGGCTATCTTATATACAAATTCGTAAAGGGCCAAATAAAGTAGGGCTTGCAGGGCT
ACCTCAGCCCCCTAGCAGATGCCGCTAAGCTTCTTACCAAAGAAATCGCAAAGCCAACAATGGCTAATTTTTCTCCTTAT
TTTTGGGCCCCCAGTATTCAGTTTTATTTAGCCCTCTTATTATGACAATTATACCCTAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATATGGTACACTTTTAGCAGGATGAGCATCAAACCTC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAA
TACTCCGCTTTCATTCATATGGTTTATAACCTGTATCGCCGAAACCAATCGG

>MLCAc

TTTTTACCTTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCCAGTATTCAGCTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCGGATACTT
TAAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>MLCAd

TTTTTCTCCTTATTTTGGGCCCCCAGTATTCAGCTTATTTTAGCCCTCTTATTATGACAATTATACCCAAGACTTTTCTCC
ACCGGATACTTTAAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATG
AGCATCAAACCTCTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCA
CTTATCTTACTCTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGAT
AACATTTCTAATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>BLMIa

TTTTTACCCTATTAGAACGAAAGGGGCTATCGTATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGC
TACCCCAACCCCTAGCAGATGCCACTAACTTCTCACCAAAGAAATCGCAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCCAGTATTCAGCTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTT
TAAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>BLMIb

GTAGGAATTGCAGGGCTACCCAGCCCCCTAGCAGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAACCAACAATG
GCTAAATTTTTCTCCTTATTTTGGGCCCCCAGTATTCAGCTTATTTTAGCCCTCTTATTATGACAATTATACCCAGGCTT
TTCTCCACCAGATACTTTAAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGC
AGGATGAGCATCAAACCTCTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTACGAAATTAG
CATAGCACTTATCTTACTCTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAG
TGTGAATAACATTTCTAATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>BLMIc

TTTTTACCCTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCCAGTATTCAGCTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTT
TAAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGAATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>KBMIa

ACGAAAGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGTCTTGCAGGGCTGCCTCAGCCCCCTAGC
AGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAAGCCAACAATGGCTAATTTTTCTCCTTATTTTGGGCCCCCAGTA
TTCAGTTTTATTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTTAAATGGGGCATCTTA
TTCTTTCTTTGTGTTTCAAGATTAAACGTATATGGTACACTTTTAGCAGGATGAGCATCAAACCTCTAAATATGCCTTACT
AGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTCTTCCCTTTATTTATTA
TAACACATTTAGATACATTGAACTAAATGAAAGTCAAGAAGCAGTATGGATAACATTTCTAATGCTCCGCTTTCATT
CATGTGGTTTGTAAACCTGTATCGCCGAAACCAATCGG

>KBMIb

TTTTTACCTTATTAGAACGAAAGGGGCTATCTTATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGACT
GCCTCAGCCCCCTAGCAGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAACCAACAATGGCTAATTTTTCTCCTTAT
TTTTTGGGCCCCCAGTATTCAGCTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGGCTAAACGTATATGGCACACTTTTAGCAGGATGAGCATCAAACCT

TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAA
TACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>KBMIc

TTTTTACCCTTATTAGAACGAAAGGGGCTATCTTATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGACT
GCCACAGCCCCTAGCAGATGCCGCTAAAGCTTCTCACCAAAGAAATCGCAAAACCAACAATGGCTAATTTTTCTCCTTAT
TTTTGGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGGCTAAACGTATATGGCACACTTTTAGCAGGATGAGCATCAAACCTC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAA
TACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>SBMA

TTTTACCCTATTAGAACGAAAGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGCT
ACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTAT
TTTTGGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCTC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAA
TACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGAGCTCCCTTTGATT

>LLMAa

TTTTACCCTATTAGAACGAAAGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGCT
ACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTAT
TTTTGGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCTC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAA
TACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGAGCTCCCTTTGATT

>LLMAb

TTTTACCCTATTAGAACGAAAGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGCT
ACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTAT
TTTTGGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCTC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAA
TACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGAGCTCCCTTTGATT

>SRNY

TTTTACCCTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGCT
ACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTAT
TTTTGGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCTC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTAA
TACTTCCGCTTTCATTTATATGATTTGTAACCTGTATCGCCGAAACCAATCGAGCTCCCTTTGATT

>CLCAa

TTTTTACCCTATTAGAACGAAAGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCGGATACTT
TAAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>PLMIa

TTTTTACCCTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGC
TACCCAGCCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTT
TAAATGAGGCATCTTATCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGAATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACCTGTATCGCCGAAACCAATCGA

>TRINa

TTTTTACCCTATTAGAACGAAAGGGGCTATCTTATATGCAGATTTCGTAAAGGACCAAATAAAGTAGGTCTTGCAGGGCT

ACCTCAGCCCCTAGCAGATGCCGCTAAGCTTCTTACCAAAGAAATCGCAAAACCAACAATGGCTAATTTTTCTCCTTAT
TTTTGAGCCCCAGTATTTAGTTTTATTTTAGCCCTTTTATTATGACAATTATACCCTAGACTTTTCTCCACCGGGTACTTT
AAATGGGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATATGGTACACTTTTAGCAGGATGGGCATCAAACCTC
TAAATATGCCTTATTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATCAGCATAGCACTTATTCTCCTT
TTCCCTATTATTATAACCACATTCAGATACATTGAACTAAATGAAAGACAAGAAGCAGTGTGAATAACATTTCTAA
TGCTTCCACTTTCATTTATATGATTTGTAACCTGTATTGCCGAGACCAACCGA

>LHMla

TTTTTACCCTATTAGAACGAAAGGGGCTATCCTATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGTCTTGCAGGGCT
GCCTCAGCCCCTAGCAGATGCCGCTAAGCTTCTCACCAAAGAAATCGCAAAGCCAACAATGGCTAATTTTTCTCCTTAT
TTTTGGGCCCCAGTATTCAGTTTTATTTTAGCCCTCTTATTATGACAATTATATCCTAGACTTTTCTCCACCAGATACTTT
AAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGATTAAACGTATATGGTACACTTTTAGCAGGATGAGCATCAAACCTC
TAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAGATTAGCATAGCACTTATCTTACTC
TTCCCTTTATTATTATAACCACATTCAGATACATTGAACTAAATGAAAGTCAAGAAGCAGTATGGATAACATTTCTAA
TGCTCCCGCTTTCATTCATGTGATTTGTAACTGTATCGCCGAAACCAATCGG

>SRMla

TTTTTACCCTATTAGAACGAAAGGGGCTATCATATATGCAAACCTCGTAAAGGACCAAATAAAGTAGGAATTGCAGGGC
TACCCAGCCCCTAGCAGATGCCGCTAAACTTCTCACCAAAGAAATCGCAAACCAACAATGGCCAATTTTTCTCCTTA
TTTTTGGGCCCCAGTATTCAGCTTTATTTTAGCCCTCTTATTATGACAATTATACCCAGACTTTTCTCCACCAGATACTT
TAAATGAGGCATCTTATTCTTTCTTTGTGTTTCAAGACTAAACGTATACGGCACACTTTTAGCAGGATGAGCATCAAACCT
CTAAATATGCCTTACTAGGAAGCCTACGGGCCATCGCACAAACAATCTCCTATGAAATTAGCATAGCACTTATCTTACT
CTTCCCTTTATTATTATGACCACATTCAGATACATTGAACTAAATGAAAGCCAAGAAGCAGTGTGGATAACATTTCTA
ATACTTCCGCTTTCATTCATATGATTTGTAACTGTATCGCCGAAACCAATCGA

APPENDIX E

LIST OF SEQUENCED POPULATIONS.

Location	code	COI	NDI	Haplo	OS	#h
Harriet Lake, MN	HLMN	5	4	B	pac	2
Isle Royale (Daisy Farm), Lake Superior, MI	IRMI	4	4	B	com	1
Niagara River, Grand Island, NY	GINY	4	4	A/B	us	1
Mud Lake, OH	MLOH	4	4	A/B		2
Long Lake, OH	LLOH	3	0	A/?		2
Lake Ontario, Waupoos, Ontario, Canada	LOCA	3	3	A		2
Moir Lake, Ontario, Canada	MLCA	3	4	A	hers	1
Camden Lake, Ontario, Canada	CLCA	3	1	A		2
Long Lake, MI	LLMI	3	3	A/B		2
Blue Lake, MI	BLMI	4	3	A/B		4
Keweenaw Bay, Lake Superior, MI	KBMI	3	3		pac/com	n/a
Limestone Lake, MN	LLMN	2	0		com	n/a
Portage Lake, MI	PLMI	1	1	A/B		1
Tippicanoe River, Tippicanoe State Park, IN	TRIN	1	1	OS	us	n/a
Harbor Beach, Lake Huron, MI	LHMI	0	1	OS	com	n/a
Saint Clair River, MI	SRMI	0	1	?/B		n/a
Stockbridge Bowl, MA	SBMA	3	1	B		1
Laurel Lake, MA	LLMA	2	2	B		1
Saw Kill, NY	SRNY	2	1	B		1
Young Lakes, NY	YLYN	4	0	A/?		1

(HAPLO= COMMON HAPLOTYPE; OS= OTHER SPECIES; #H= NUMBER OF HAPLOTYPES WITHIN THE *M. LUSTRICA* CLADE; PAC=*M. PACHYTA*; COM= *M. COMALENSIS*; HERS= *M. HERSHLERI*; US= UNIDENTIFIED *MARSTONIA*)

APPENDIX F

EXTRACTION PROTOCOL.

Extraction Materials

1. 2X CTAB: 0.1M Tris-HCl pH 8.0, 1.4M NaCl, 0.02M EDTA, 2% CTAB, 0.2% mercaptoethanol (100 ml: 10 ml 1M Tris pH 8.0, 28 ml 5M NaCl, 4 ml 0.5M EDTA pH 8.0, 20 ml 10% CTAB, 0.2 ml mercaptoethanol, 37.8 ml autoclaved ultra-pure water)
 - 1M Tris-HCl, pH8.0: dissolve 121.1g of Tris base in 800ml of ultra-pure water. Adjust the pH to 8.0 by adding concentrated HCl (approximately 42ml). Allow the solution to cool to room temperature before making final adjustments to the pH. Adjust the volume of the solution to 1L with ultra-pure water. Dispense into aliquots and sterilize by autoclaving 15 minutes in liquid cycle.
 - 0.5M EDTA, pH8.0: dissolve 186.1g of EDTA to 800ml of ultra-pure water. Adjust the pH to 8.0 with NaOH (approximately 20 g of NaOH pellets). Dispense into aliquots and sterilize by autoclaving 15 minutes in liquid cycle.
 - 5M NaCl: dissolve 292.2g of NaCl in 800 ml of ultra-pure water. Adjust the volume to 1 liter with ultra-pure water. Dispense into aliquots and sterilize by autoclaving.
 - 10% CTAB: 10g CTAB in 100ml autoclaved ultra-pure water.
2. Proteinase K: 20 mg proteinase K in 1 ml of autoclaved ultra-pure water.
3. Chloroform/Isoamyl Alcohol (24:1): 96ml Chloroform, 4ml Isoamyl Alcohol.
4. Isopropanol
5. TE buffer: 10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0 (50ml: 0.5ml 1M Tris-HCl pH 8.0, 0.1ml 0.5M EDTA pH 8.0, 49.4ml autoclaved ultra-pure water).

Procedures

1. Label microcentrifuge tubes appropriately.
2. Transfer specimen into a labeled microcentrifuge tube containing 100 μ l 2X CTAB.
3. Grind specimen with a pestle.
4. Add 600 μ l 2X CTAB.
5. Incubate at 65 $^{\circ}$ C for one hour.
6. Take specimens out of 65 $^{\circ}$ C heat block and let them cool to room temperature.
7. Add 5 μ l Proteinase K (20mg/ml).
8. Incubate at 37 $^{\circ}$ C for 3 hours or longer. (**NOTE:** I usually do it overnight)
9. Add equal volume of chloroform: isoamyl alcohol (24:1).
10. Mix well by inversion 200 times.
11. Centrifuge at 10,000 rpm for 10 minutes.
12. Move the supernatant into a new microcentrifuge tube
 - Leave the lower layer in the microcentrifuge tube with lid open in the hood.
 - When all the chloroform evaporated from the microcentrifuge tube, discard the microcentrifuge tube into the trash can
14. Precipitate DNA with 600 μ l isopropanol. Mix well and store at -20 $^{\circ}$ C overnight.
15. Centrifuge at 14,000 rpm for 30 minutes
16. Discard the upper isopropanol layer
17. Add 200 μ l 70% ethanol
18. Mix well by inverting the tubes several times
19. Centrifuge at 14,000 rpm for 5 minutes
20. Discard the upper layer (70% ethanol)
 - It is important to discard the 70% ethanol immediately after centrifugation. DNA pellet is loose in 70% ethanol.
21. Invert the microcentrifuge tube and let the DNA pellet air dry for about an hour or longer.
22. Resuspend the DNA pellet in TE buffer (25-50 μ l).
 - This may take some time (up to several hours) since the DNA is of high molecular weight.
23. Incubate at 37C for several hours. Store DNA at 4C.
 - Purified DNA is generally stored (at approximately 4 $^{\circ}$ C) during times of active use.

APPENDIX G

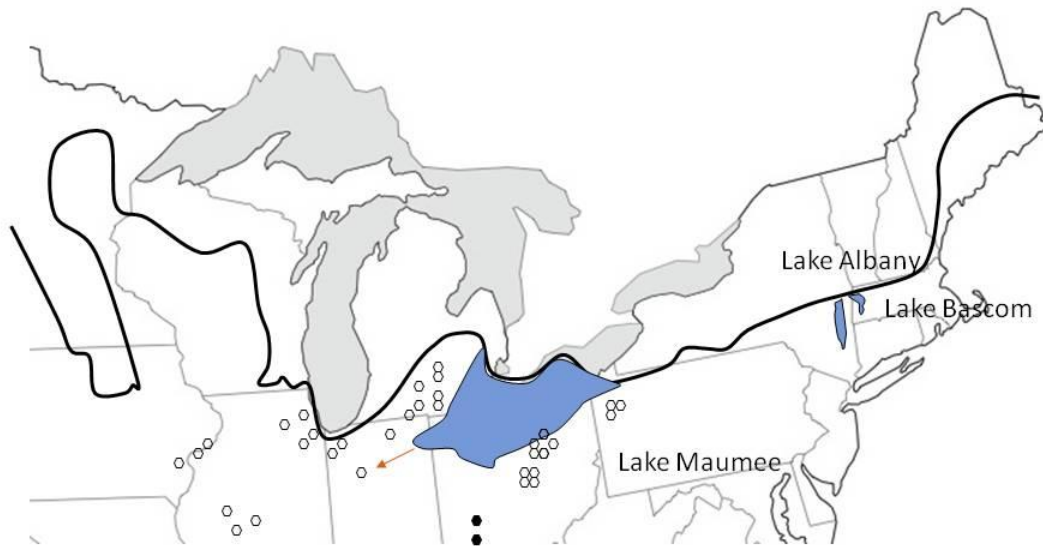
GLACIAL MAPS

This series of idealized maps illustrate glacial retreat patterns and pro-glacial lakes of the Wisconsinian from 18,000 kya to the present. Empty dots indicate historical populations (Thompson 1977) and filled in dots are hypothesized established and isolated populations for the given time period.

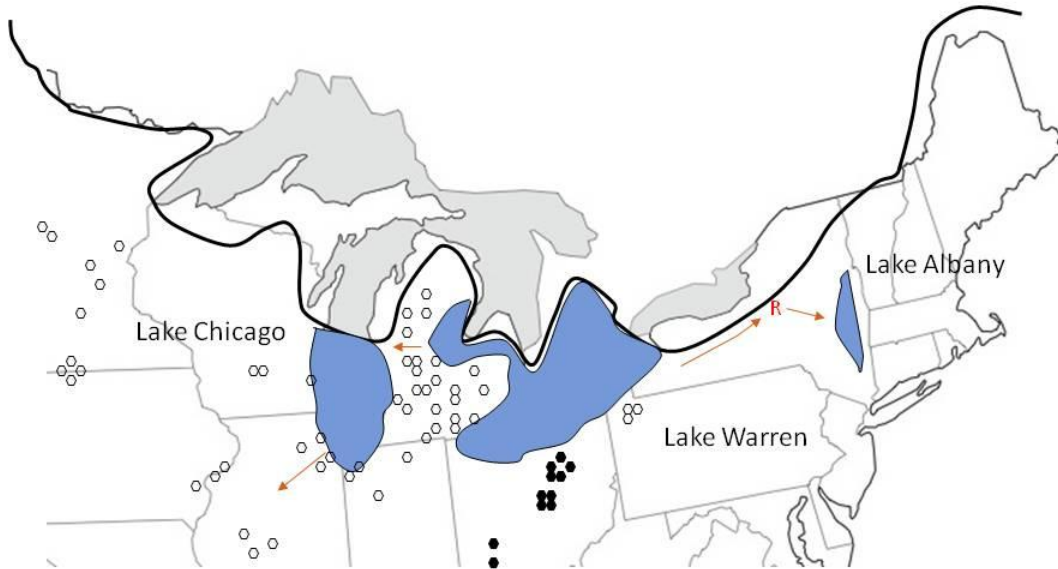
18,000 BP



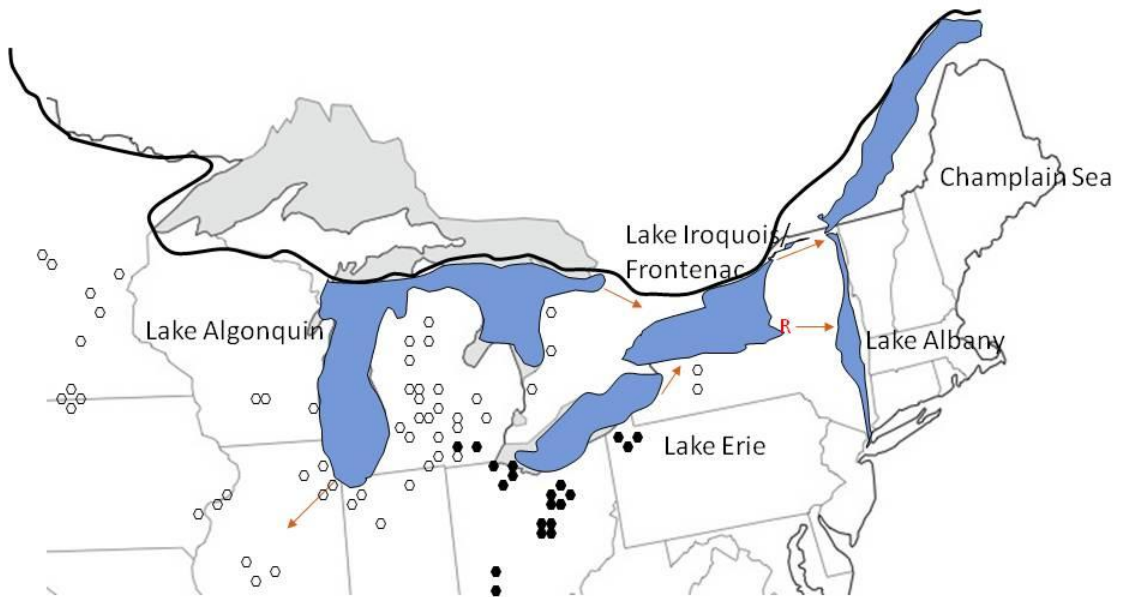
14,000 BP



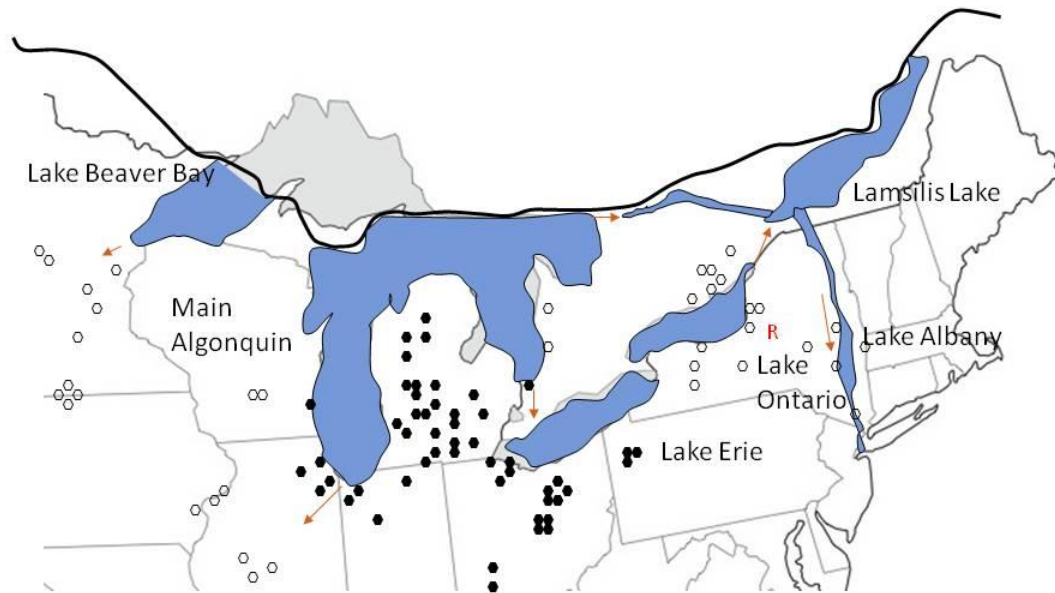
12,000 BP



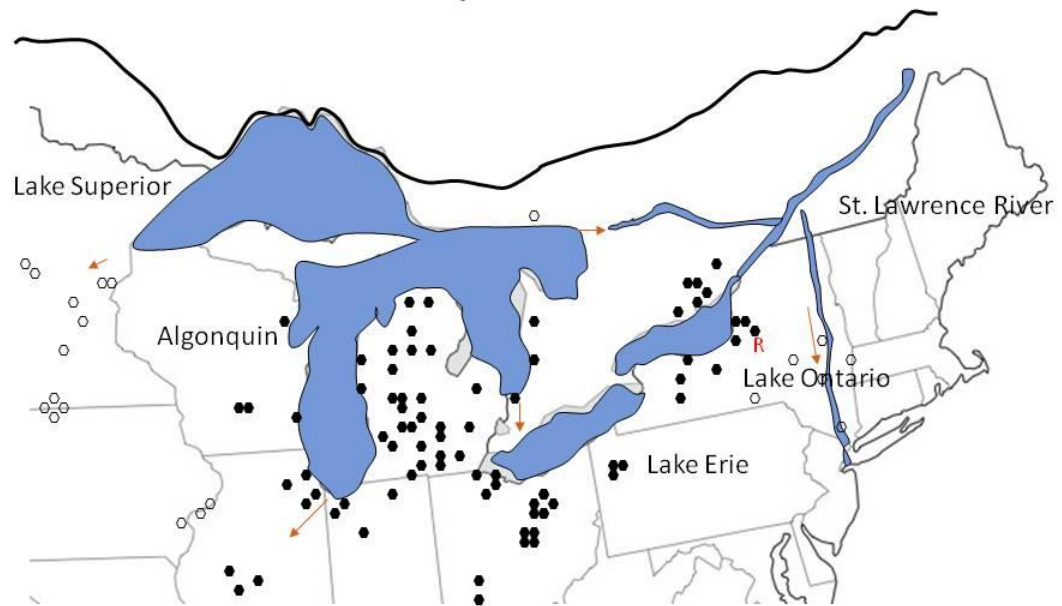
11,000 BP



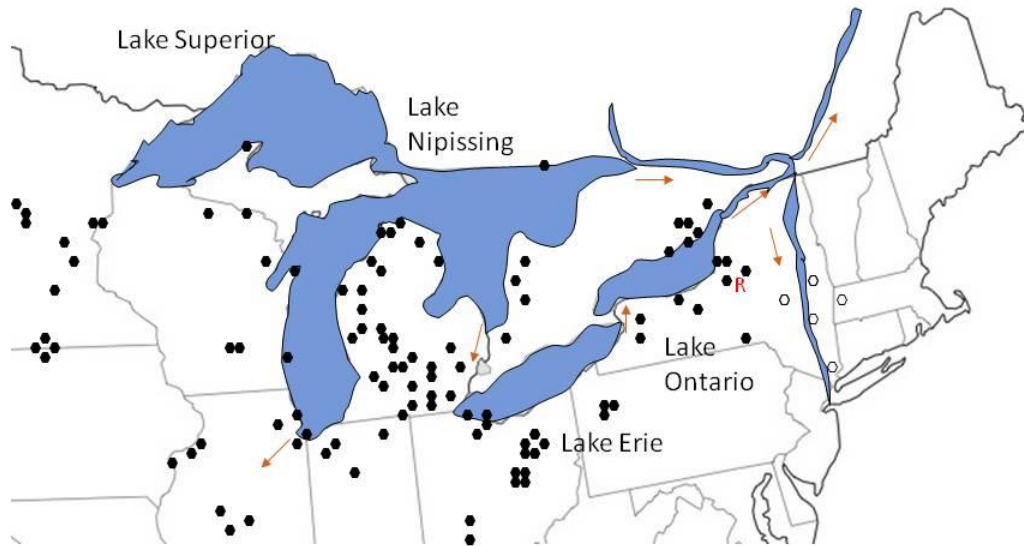
10,000 BP



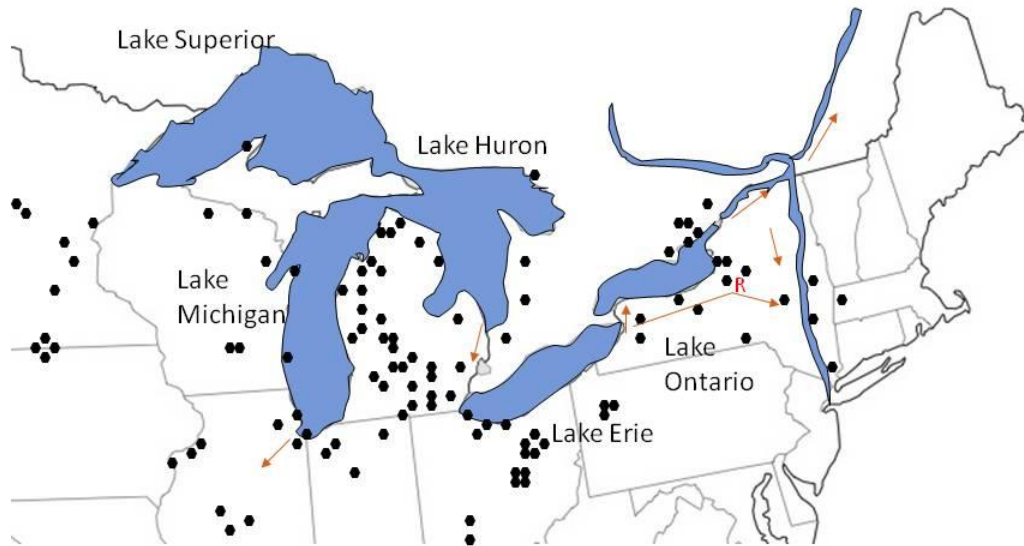
8,000 BP



6,000 BP



Present - 4,000 BP



APPENDIX H

NUCLEOTIDE SUBSTITUTION PATTERNS

Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution for COI [1]

	A	T	C	G
A	-	<i>2.71</i>	<i>2.06</i>	24.43
T	<i>2.16</i>	-	12.09	<i>1.75</i>
C	<i>2.16</i>	15.91	-	<i>1.75</i>
G	30.22	<i>2.71</i>	<i>2.06</i>	-

NOTE: Each entry shows the probability of substitution from one base (row) to another base (column) instantaneously. Only entries within a row should be compared. Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics. The nucleotide frequencies are 0.249 (A), 0.312 (T/U), 0.237 (C), and 0.201 (G). The transition/transversion rate ratios are $k1 = 13.989$ (purines) and $k2 = 5.874$ (pyrimidines). The overall transition/transversion bias is $R = 4.495$, where $R = [A * G * k1 + T * C * k2] / [(A + G) * (T + C)]$. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 658 positions in the final dataset. All calculations were conducted in MEGA4 [2].

1. Tamura K, Nei M & Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences (USA)* 101:11030-11035.
2. Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599.

Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution for NDI [1]

	A	T	C	G
A	-	<i>1.45</i>	<i>1.11</i>	15.49
T	<i>1.44</i>	-	18.7	<i>0.69</i>
C	<i>1.44</i>	24.33	-	<i>0.69</i>
G	32.1	<i>1.45</i>	<i>1.11</i>	-

NOTE: Each entry shows the probability of substitution from one base (row) to another base (column) instantaneously. Only entries within a row should be compared. Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics. The nucleotide frequencies are 0.306 (A), 0.309 (T/U), 0.237 (C), and 0.148 (G). The transition/transversion rate ratios are $k1 = 22.336$ (purines) and $k2 = 16.788$ (pyrimidines). The overall transition/transversion bias is $R = 7.301$, where $R = [A * G * k1 + T * C * k2] / [(A + G) * (T + C)]$. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 544 positions in the final dataset. All calculations were conducted in MEGA4 [2].

HAPLOTYPE TABLE FOR COI.

1122233334479911111111111111111222222222222222233333333334444444444
58157016925869012222344455667778990012223455566778899011355689902334566777
57036921240325147358474256492484092514628937380180281328247

Ref	GAAGACCGCCGCCGGATCTGTGTTATGCGCCATTAGCCCAAAATGACAACATCCCATCTAAAAAGACATTCTA									
HLMNa 3									
HLMNc	...T...	AATA	...	GG	1		
HLMNd	...T...	AAGTA	...	GG	1		
IRMIa	A...T	A...T	T.C	..GGGG	4	
IRMId	...T...	AGTA	...	GGG	1	
GINYa	...TT...	AAAG	..T	C	AT	..G	..CTG	..M.....GG.TAG 1
GINYb	...T...	ATA	...	GG	19		
GINYd	...TT...	AAAG	..T	C	AT	..G	..CTGG.TAG 1
MLOHd	...T...	AG	..TA	...	GG	1	
LLMIc	...T...	ATA	...	GGTG	1		
LOCAa	...T...	AA	TA	...	GG	2	
MLCAb	..G	ATT	..A	..TC	..G	..GGG	1
BLMIa	...T...	ATA	...	TGG	1		
BLMIb	...T...	AC	A	..G	..TA	...	GG
BLMIc	...T...	ATA	...	C	..GG	1	
KBMIdG.....G.... 1									
KBMId	...T...G.....G.... 1								
CLCAa	...T...	ATA	...	GG	1		
LLOHb	...T...	AAG	..TA	...	GCG	1
LLOHc	...T...	ATA	...	G	----- 1			
LLMNa	...T	AT	..G	..T	AT	..GGG
TRINa	-----	AAAG	..T	C	AT	..G	..CTGG.TAG 1
SBMAa	...T...	ATA	...	G	..GG	5	
SRNYa	...T...	ATA	...	GG	2		

APPENDIX J

HAPLOTYPE TABLE FOR NDI

Collapsed sequences:

HLMNa with HLMNb
 HLMNc with HLMNd
 HLMNc with IRMIa
 HLMNc with GINYb
 HLMNc with GINYc
 HLMNc with MLOHa
 HLMNc with MLOHc
 HLMNc with LLMiC
 HLMNc with BLMiA
 HLMNc with BLMiC
 HLMNc with SBMA
 HLMNc with LLMAa
 HLMNc with LLMAb
 HLMNc with SRNY
 HLMNc with PLMiA
 HLMNc with SRMiA
 IRMiA with IRMiB
 IRMiA with IRMiC
 IRMiA with KBMiA
 GINYa with GINYd
 GINYa with TRINa
 LLMiA with LLMiB
 LOCAb with MLCaA
 LOCAb with MLCaC
 LOCAb with CLCAa
 KBMiB with KBMiC

123333466778889111111111111111122222222222222

8103698692845790223444566888899991223444555578

0390147928347925689581679258905

Ref GCCCTTCCCACCAAAGTACTCAAAACTCAACCACTATCTACCTTGC

HLMNa 2
 HLMNcC..... 16
 IRMiA ...T...T....G...T.T.....A. 4
 GINYa A.TT..T.T...GGG....T..G...T.....CTCCTTCCA. 3
 MLOHbT.....C..... 1
 MLOHdC....T..... 1
 LLMiAC.....T..... 2
 LOCAaA...C.....T 1
 LOCAbG.....C..... 4
 LOCAcC.....T 1
 MLCaB ...T...T.....T.....A. 1
 MLCAdA..G.....C..... 1
 BLMiBG.....C.....C..... 1
 KBMiBG..... 2
 LHMiA ...T...TT.....T.T.....G.....A. 1
 hershl ...T.....T.G...C....G.....A. 1
 halcyo ATTTCTT.T.TTGG..C....G.GG..GGA.....C.....A. 1
 agarhe .TTTCCT...TTGG..C....G.GG...G.AT.....C.....A. 1

LITERATURE CITED

- Baker, F.C. 1928. The freshwater Mollusca of Wisconsin. Part 1. Gastropoda. *Wisconsin Geological and Natural History Survey Bulletin* 70: 1-494.
- Bandelt H-J, P. Forster, and A. Röhl (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48.
- Barnett, P.J. 1992. Quaternary Geology of Ontario. In *Geology of Ontario*, eds. P.C. Thurston, H.R. Williams, R.H. Sutcliffe, and G.M. Stott, Ontario Geological Survey Special Volume 4:1011–1088.
- Benn, D. I and D. J. Evans. 1998. *Glaciers and Glaciation*. Arnold, Great Britain.
- Berry, E. G. 1943. The Amnicolidae of Michigan: distribution, ecology, and taxonomy. *Miscellaneous Publications, Museum of Zoology, University of Michigan* No. 57.
- Bierman P. R. and D. P. Dethier. 1986. Lake Bascom and the deglaciation of northeastern Massachusetts. *Northeastern Geology* 8 (1/2): 32-43.
- Boag, D. A. 1986. Dispersal in pond snails: potential role of waterfowl. *Can. J. Zool.* 64: 904-909.
- Bryson, R. A., W. M. Wendland, J. D. Ives, and J. T. Andrews. 1969. Radiocarbon Isochrones on the Disintegration of the Laurentide Ice Sheet. *Arctic and Alpine Research* 1 (1): 1-14.
- Burch, J. B. 1980. Species list, ranges and illustrations. *Walkerana* 1(3): 81-215.
- Burch, J. B. 1982. *North American Freshwater Snails: Identification Keys, Generic Synonymy, Supplemental Notes, Glossary, References, Index*. pp. 217-337. Ann Arbor, MI: Museum of Zoology, The University of Michigan.
- Burch, J. B. 1988. *North American Freshwater Snails: Introduction, Systematics, Nomenclature, Identification, Morphology, Habitats, Distribution*. Ann Arbor, MI: Museum of Zoology, The University of Michigan.
- Burch, J. B. 1989. Identification and morphology, habitats and distribution. *Walkerana* 2 (6): 25-80.
- Campbell, D. 2006. Cladistics and Molecular Techniques: A Primer. In: *The Mollusks: A Guide to Their Study, Collection, and Preservation*. Sturm et al (Eds.), American Malacological Society.

- Cao, Y., D. D. Williams, and N. Williams. 1998. How important are rare species in aquatic community ecology and bioassessment? *Limnology and Oceanography*, 43 (7): 1403-1409.
- Coote, T. and D. Roeder, 1999. *The Relative Abundance of Pyrgulopsis lustrica and its Movements in Stockbridge Bowl, September-December, 1999*. Great Barrington, MA.
- Coote, T. and D. Strayer. 2008. Gastropods of the Hudson River shoreline: subtidal, intertidal, and upland communities. In: *Final Reports of the Tibor T. Polgar Fellowship Program*. Hudson River Foundation.
- Curry, R. A. 2007. Late glacial impacts on dispersal and colonization of Atlantic Canada and Maine by freshwater fishes. *Quaternary Research*, doi: 10.1016/j.yqres.2006.11.002.
- Davis, G. M. and M. Mazurkiewicz. 1985. Systematics of *Cincinnatia winkleyi* (Gastropoda: Hydrobiidae). *Proceedings of the Academy of Natural Sciences of Philadelphia* 137: 28-47.
- Davis, G.M., C.E. Chen, C. Wu, T.F. Kuang, X. G. Xing, L. Li, W.J. Liu, and Y.L. Yan. 1992. The Pomatiopsidae of Hunan, China (Gastropoda: Rissoacea). *Malacologia* 34:143-342.
- DeChaine, E.G. and A.P. Martin. 2004. Historic cycles of fragmentation and expansion in *Parnassius smintheus* (Papilionidae) inferred using mitochondrial DNA. *Evolution* 58 (1): 113-127.
- Dillon, R. T. and R. C. Frankis. 2004. High levels of mitochondrial DNA sequence divergence in isolated populations of freshwater snails of the genus *Goniobasis* Lea, 1862. *American Malacological Bulletin* 19: 69-77.
- Eschman, D. F. 1985. Summary of the Quaternary History of Michigan, Ohio, and Indiana. *Journal of Geological Education* 33: 161-167.
- Emerson, B. C. and G. M Hewitt. 2005. Phylogeography. *Current Biology* 15 (10); 367-71.
- ENSR. 1998. *Determination of the Presence and Distribution of the Two Protected Snail Species, Marstonia lustrica and Valvata sincera, in Stockbridge Bowl, Lenox and Stockbridge, MA, and Relation to Lake Management*. Northborough, MA.
- ESRI. 2009. www.esri.com

- Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
- Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.
- Fluxus Engineering 2010. Network 4.5.6. www.fluxus-engineering.com
- Fullerton, D.S., R.B. Colton, and C.A. Bush. 2004. Limits of mountain and continental glaciations east of the continental divide in northern Montana and north-western North Dakota, U.S.A. In: *Quaternary Glaciations: Extent and Chronology, Vol 2*. Ehlers J. and P.L. Gibbard eds. Elsevier.
- Fugro 1996. *Lake and Watershed Management Plan for Stockbridge Bowl, Stockbridge, Massachusetts*. Northborough, MA.
- Hall, B. G. 2008. *Phylogenetic Trees Made Easy*, 3rd ed. Sinauer, Sunderland.
- Harman, W. and C. Berg. 1971. The freshwater snails of central New York. *Search* (Cornell University Agricultural Experiment Station, Ithaca, New York) 1 (4): 1-68.
- Hershler, R. 1994. A review of the North American freshwater snail genus *Marstonia* (Hydrobiidae). *Smithsonian Contributions to Zoology*, 554: 1-115.
- Hershler, R., H-P Liu, and F. G. Thompson. 2003. Phylogenetic relationships of North American nymphophiline gastropods based on mitochondrial DNA sequences. *Zoologica Scripta* 32 (4): 357-366.
- Hershler, R. and H-P Liu. 2004a. A molecular phylogeny of aquatic gastropods provides a new perspective on biogeographic history of the Snake River region. *Molecular Phylogenetics and Evolution* 32: 927-937.
- Hershler, R. and H-P Liu. 2004b. Taxonomic reappraisal of species assigned to the North American freshwater gastropod subgenus *Natricola* (Rissooidea: Hydrobiidae). *Veliger* 47: 66-81.
- Hershler, R., H-P Liu, T.J Frest, E.J. Johannes, and W.H. Clark. 2006. Genetic structure of the western North American aquatic gastropod genus *Taylorconcha* and description of a second species. *J. Molluscan Studies* 72: 167-177.
- Hershler, R., H-P Liu, T.J Frest, E.J. Johannes. 2007a. Extensive diversification of pebblesnails (Lithoglyphidae: *Fluminicola*) in the upper Sacramento River Basin, northwestern USA. *Zoological J. Linnean Society* 149: 371-422.

- Hershler, R., H-P Liu, and D.W. Sada. 2007b. Origin and diversification of the Soldier Meadow springsnails (Hydrobiidae: *Pyrgulopsis*), a species flock in the northwestern Great Basin, United States. *J. Molluscan Studies* 73: 167-183.
- Hershler, R., H-P Liu, and D.L. Gustafson. 2008. A second species of *Pyrgulopsis* (Hydrobiidae) from the Missouri River Basin, with molecular evidence supporting faunal origin through Pliocene stream capture across the northern continental divide. *J. Molluscan Studies* 74: 403-413.
- Hershler, R. and H-P Liu. 2011. ZooKeys 77: 1–16 (2011). doi: 10.3897/zookeys.77.935.
- Hershler, R. and W. F. Ponder. 1998. A Review of Morphological Characters of Hydrobioid Snails. *Smithsonian Contributions to Zoology* 600:1-55
- Hershler, R. and F.G. Thompson. 1987. North American Hydrobiidae (Gastropoda: Rissoacea): redescription and systematic relationships of *Tyronia* Stimpson, 1865 and *Pyrgulopsis* Call and Pilsbry, 1886. *The Nautilus* 101 (1): 25-32.
- Hewitt, G. M. 1995. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247-276.
- Hewitt, G. M. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405, June 22.
- Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Phil. Trans. Royale Society, London*. 359: 183-195. Doi: 10.1098/rstb.2003.1388.
- Holman, J. A. 1992. Late quaternary herpetofauna of the central Great Lakes region, U.S.A.: zoogeographical and paleoecological implications. *Quaternary Science Reviews* 11: 345-351.
- Horst, T. J. and R. R. Costa. 1975. Seasonal migration and density patterns of the freshwater snail *Amnicola limosa*. *Nautilus* 89:56-59.
- IUCN 2004. *The IUCN Red List of Threatened Species*. www.redlist.org.
- Jokinen, E. H. 1983. The freshwater snails of Connecticut. *State Geological and Natural History Survey of Connecticut. Bulletin* 109:36.
- Jokinen, E. H. 1992. The freshwater snails (Mollusca: Gastropoda) of New York State. *New York State Museum Bulletin* 482.
- Jokinen, E. H. and Biological Survey-New York State Museum. 1987. Structure of freshwater snail communities: species-area relationships and incidence categories. *American Malacological Bulletin* 5: 9-19

- Jokinen, E. H. and J. Pondick. 1981. Rare and endangered species: Freshwater gastropods of southern New England. *Bull. Amer. Malacol. Union*, 1981:52-53.
- Kabat, A. and R. Hershler. 1993. The prosobranch snail family Hydrobiidae (Gastropoda:Rissooidea): Review of classification and supraspecific taxa. *Smithsonian Contributions to Zoology* 547: 1-94.
- Kelley, D.W., H.J. MacIsaac, and D.D. Heath. 2006. Vicariance and dispersal effects on phylogeographic structure and speciation in a widespread estuarine invertebrate. *Evolution* 60 (2): 257-267.
- Kesler, D. H. 1980. Seasonal abundance of *Amnicola limosa* (Hydrobiidae) eggs and individuals in a Rhode Island pond. *Nautilus* 94: 25-26.
- Knowles, L.L., D.J. Futuyma, W.F. Eanes, and B. Rannala. 1999. Insight into speciation from historical demography in the phytophagous beetle genus *Ophraella*. *Evolution* 53 (6): 1846-1856.
- Koteff, C. and F. Pessl. 1981. *Systematic Ice Retreat in New England*. U.S. Dept. of the Interior, Geological Survey.
- Larson, G. and R. Schaetzl. 2001. Review: Origin and evolution of the Great Lakes. *J. Great Lakes Res.* 27 (4): 518-546.
- Lewis, J. 1856. Shellbearing species of mollusca observed in portions of Herkimer and Otswego counties, New York. *Proc. of the Boston Socceity of Natural History* 6: 2-4.
- Lewis, J. 1860. Catalogue of the Mollusks in the vicinity of Mohawk, New York. *Proc. of the Academy of Natural Sciences of Philadelphia* ?:17-19.
- Lewis, J. 1868a. Remarks on some of the mollusca of the valley of the Mohawk. *Amer. J. of Conchology* 4: 241-245.
- Lewis, J. 1868b. Notes on certain fresh-water shells, observed in the vicinity of Mohawk, N.Y. *Amer. J. of Conchology* 4: 2-4.
- Lewis, J. 1874. Land and fresh water shells of the State of New York. *Bulletin of the Buffalo Academy of Natural Science* 2: 127-142.
- Lewis, D. B. and J. J. Magnuson. 2000. Landscape spatial patterns in freshwater snail assemblages across northern highland catchments. *Freshwater Biology* 43 (3), 409-420.
- Lewis, R. 1997. <http://www.conncoll.edu/ccrec/greenet/arbo/publications/34/chp1.htm>.

- Linsley, 1845. Amer. J. Soi. In: Smith, S. and T. Prime. 1870. Report on the mollusca of Long Island, N.Y., and of its dependencies. *Annals of the Lyceum of Natural History of New York* 9: 377-407.
- Liu, H-P, R. Hershler, and K. Clift. 2003. Mitochondrial DNA sequences reveal extensive cryptic diversity within a western American springsnail. *Mol. Ecology* 12: 2771-2782.
- Liu, H-P and R. Hershler. 2005. Molecular systematic and radiation of western North American nymphophiline gastropods. *Molecular Phylogenetics and Evolution* 34: 284-298.
- Liu, H-P and R. Hershler. 2007. A test of the vicariance hypothesis of western North American freshwater biogeography. *J. Biogeogr.* 34: 534-548.
- Lodge, D. M., K. M. Brown, S. P. Klosiewski, R. A. Stein, A. P. Covich, B. K. Leathers, and C. Brönmark. 1987. Distribution of freshwater snails: Spatial scale and the relative importance of physicochemical and biotic factors. *American Malacological Bulletin* 5 (1): 73-84.
- Ludlam, S., K. S. Hutchinson, and G. E. Henderson. 1973. *The Limnology of Stockbridge Bowl, Stockbridge, Massachusetts*. University of Massachusetts Water Resources Center, Completion Report FY-73-4. 59 p.
- Lydeard, C., R. H. Cowie, W. F. Ponder, A. E. Bogan, P. Bouchet, S. A. Clark, K. S. Cummings, T. J. Frest, O. Gargominy, D. G. Herbert, R. Hershler, K. E. Perez, B. Roth, M. Seddon, E. E. Strong, and F. G. Thompson. 2004. The global decline of nonmarine mollusks. *BioScience* 54 (4): 321-30.
- MacNamara, M.C. and W.N. Harman. 1975. Further studies of the mollusca of the Otsego Lake area. *Nautilus* 89 (3): 87-90.
- Manel, S., M. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18 (4): 189-197.
- Massachusetts Natural Heritage and Endangered Species Program. 2010. www.state.ma.us/dfwele/dfw/nhesm.
- Mayewski, P. A., G. H. Denton, and T. J. Hughes. 1981. *Late Wisconsin Ice Sheets in North America*. In *The Last Great Ice Sheets*. Denton and Hughes eds., John Wiley and Sons, New York.
- McLain, D. 2003. *Status of 4 State-listed Snails in Western Massachusetts in 2002*. Report to the Massachusetts Natural Heritage and Endangered Species Program. Easthampton, MA.

National Atlas. 2010.

http://www.nationalatlas.gov/articles/geology/a_glacial.html

Økland, Jan. 1990. *Lakes and Snails: Environment and Gastropoda in 1,500 Norwegian lakes, ponds and rivers*. Universal Book Services.

Perez, K.E., W.F. Ponder, D.J. Colgan, S.A. Clark, and C. Ledyard. 2005. Molecular phylogeny and biogeography of spring-associated hydrodiid snails of the Great Artesian Basin, Australia. *Mol. Phyl. And Evolution* 34: 545-556.

Perez, K.E. and R.L Minton. 2008. Practical applications for systematic and taxonomy in North American freshwater gastropod conservation. *J. N.A. Benthological Soc.* 27 (2): 471-483.

Pilsbry, H. A. 1890. Preliminary notices of new Amnicolidae. *Nautilus* 4: 52-53.

Rittenour, T.M. 1997.

<http://www.bio.umass.edu/biology/conn.river/hitchcock.html>.

Roeder, D. and T. Coote, 2000. *The Distribution of Pyrgulopsis lustrica in Western Massachusetts, June-August 2000*. Report to The Massachusetts Natural Heritage Program. Great Barrington, MA.

Rowe, K.C., E.J. Heske, P.B. Brown, and K.N. Paige. 2004. Surviving the ice: northern refugia and postglacial colonization. *PNAS* 101 (28): 10355-10359.

Schaetzl, R.J., F.J. Krist, P.R. Rindfleisch, J. Liebens, and T.E. Williams. 2000. Postglacial landscape evolution of northeastern lower Michigan, interpreted from soils and sediments. *A. Amer. Assoc. of American Geographers* 90 (3): 443-466.

Schmidt, R. E. 1986. Zoogeography of the Northern Appalachians. In: *The Zoogeography of North American Freshwater Fishes*. Hocutt and Wiley (Eds.), John Wiley & Sons.

Siegert, M.J. 2001. *Ice Sheets and Late Environmental Change*. John Wiley & Sons, Chichester.

Smith, D. G. 1995. *Keys to the Freshwater Macroinvertebrates of Massachusetts*, 2nd edition. Sunderland, MA.

Smith, D. G. 1983. The zoogeography of the freshwater mussels of the Taconic and southern Green Mountain region of northeastern North America (Mollusca: Pelecypoda: Unionacea) *Can. J. Zool.* 60: 261-267.

Statsoft. 1999. STATISTICA for Windows. Tulsa Oklahoma.

- Strayer, D. 1987. Ecology and zoogeography of the freshwater mollusks of the Hudson River basin. *Malacological Review* 20:1-68.
- Strothard, J.R., S. Hughes, and D. Rollinson. 1996. Variation within the internal transcribed spacer (ITS) of ribosomal DNA genes of intermediate snail hosts within the genus *Bulinus* (Gastropoda: Planorbidae). *Acta Tropica* 61: 19-29.
- Sugden, D.E. 1977. Reconstruction of the morphology, dynamics and thermal characteristics of the Laurentide Ice Sheet at its maximum. *Arctic and Alpine Research* 9 (1): 21-47.
- Tamura K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
- Thompson, F. G. 1977. The hydrobiid snail genus *Marstonia*. *Bulletin of the Florida State Museum, Biological Sciences* 21 (3):113-158.
- Thompson, F.G. and Hershler, R. 2002. Two genera of North American freshwater snails: *Marstonia* Baker, 1926, resurrected to generic status, and *Floridobia*, new Genus (Prosobranchia: Hydrobiidae: Nymphophilinae). *The Veliger*, 45(3): 269-271.
- United States Geological Survey. 2009. Science in Your Watershed: <http://water.usgs.gov>
- Wall, R. C. 1977. Seasonal movements of the pond snail, *Lymnaea catascopium* in a northern lake. *Nautilus* 91 (2): 47-51.
- Wilke, T. and G.M. Davis. 2000. Intraspecific mitochondrial sequence diversity in *Hydrobia ulvae* and *Hydrobia ventrosa* (Hydrobiidae: Rissooidea: Gastropoda): do their different life histories affect biogeographic patterns and gene flow? *Biol. J. Linnean Society* 70: 89-105.
- Wilke, T., G. M. Davis, A. Falniowski, F. Giusti, M. Bodon, and M. Szarowska. 2001. Molecular systematic of Hydrobiidae (Mollusca: Gastropoda: Rissooidea): testing monophyly and phylogenetic relationships. *Proc. Acad. Nat. Sci. Philadelphia* 151: 1-21.
- Wilke, T., M. Pfenninger, and G.M. Davis. 2002. Anatomical variation in cryptic mudsnail species: statistical discrimination and evolutionary significance. *Proc. Acad. Nat. Sci. Philadelphia* 152: 45-66.
- Wilke, T., R. Schultheiß, and C. Albrecht. 2009. As time goes by: a simple fool's guide to molecular clock approaches in invertebrates. *Amer. Malac. Bull.* 27: 25-45.
- Yang, J., P.F. Karrow, and G.L. Mackie. 2001. Paleoecological analysis of molluscan assemblages in two marl deposits in the Waterloo region, southwestern Ontario, Canada. *J. of Paleolimnology* 25: 313-328.